Paper 1557-2014

The Query Builder: The Swiss Army Knife of SAS® Enterprise Guide®

Jennifer First-Kluge and Steven First, Systems Seminar Consultants, Inc.

ABSTRACT

The SAS® Enterprise Guide® Query Builder is one of the most powerful components of the software. It enables a user to bring in data, join, drop and add columns, compute new columns, sort, filter data, leverage the advanced expression builder, change column attributes, and more! This presentation provides an overview of the major features of this powerful tool and how to leverage it every day.

WHAT IS THE QUERY BUILDER?

Enterprise Guide is a user interface for SAS. EG passes SAS code to a server to run the SAS code (whether you are writing code or using the tasks in EG). The server could be your local machine, a server on your network, it could even be on another platform. A task is a point and click user interface in Enterprise Guide that generates SAS code. Each task has an analogous PROC (although not 1 to 1 relationship).

One of the most powerful tasks is the Query Builder. The Query Builder allows you to manipulate data tables in a variety of ways, including joining tables, selecting variables, filtering data, sorting data, changing data sources, computing columns, setting up prompts, de-duping observations, adding titles and footnotes, limiting output, changing query options, and more.

What's SQL

A query language used by many software packages.

The Query Builder in Enterprise Guide generates PROC SQL code behind the scenes.

We don't code SQL because the query builder will do it for us.

SQL

Example of SQL code in SAS:

```
proc sql options;
  create table|view as        /* output table to create */
  select column(s)            /* Select/create columns  */
  from table-name | view-name /* Name input sources     */
  where expression            /* Sub-set rows from table*/
  group by column(s)          /* Group rows for summary */
  having expression           /* Subset GROUP BY results*/
  order by column(s)          /* Sort resulting rows    */
   ;
quit;
```

Why Should I Care About the Query Builder?

Quickly bring together data

Manipulate/modify data easily

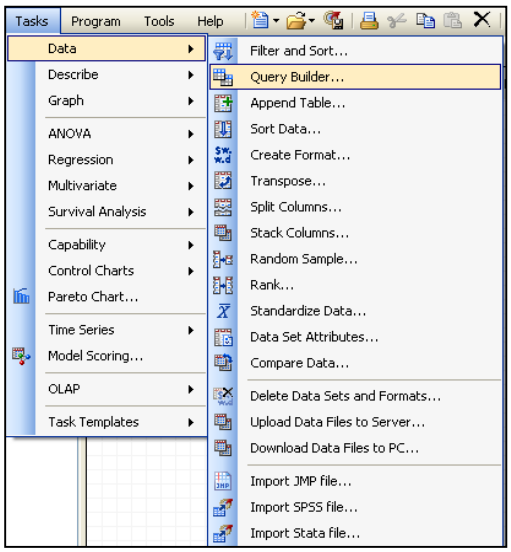
Error free queries

Quicker, less error prone that coding.

Creating a New Query

In the Process Flow or Project Window, highlight a data set.

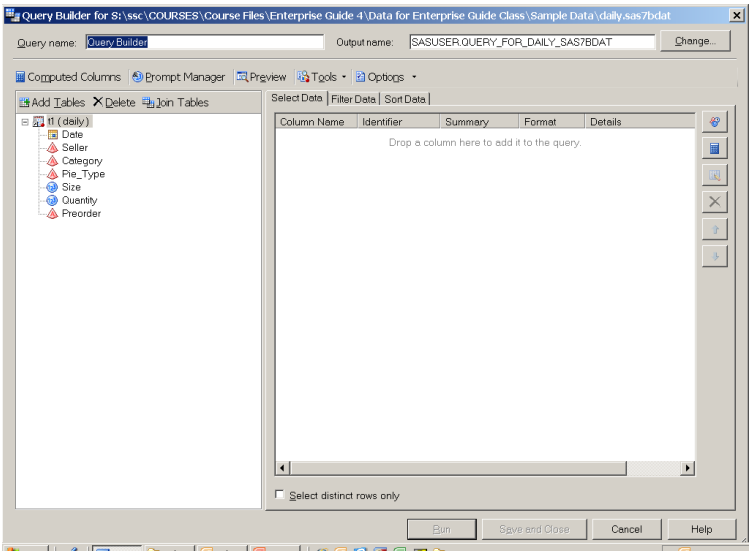
On the Data, Tasks menu, select Query Builder.



The Query Builder

The Tables list shows the columns that make up query tables.

The Select Data tab lists the columns included in the query.

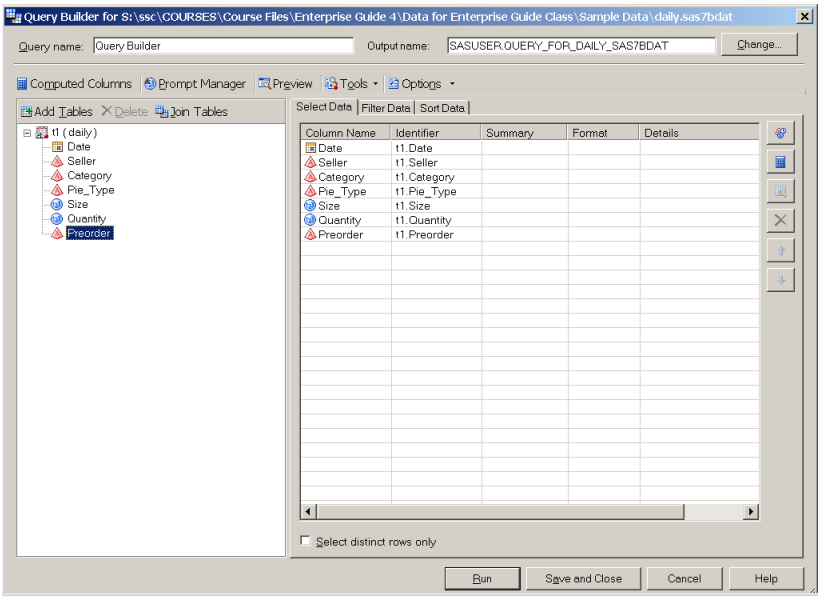


Adding Columns to a Query

Double-click columns to add them to the query.

The  and  buttons move columns up/down.

The  button deletes columns.

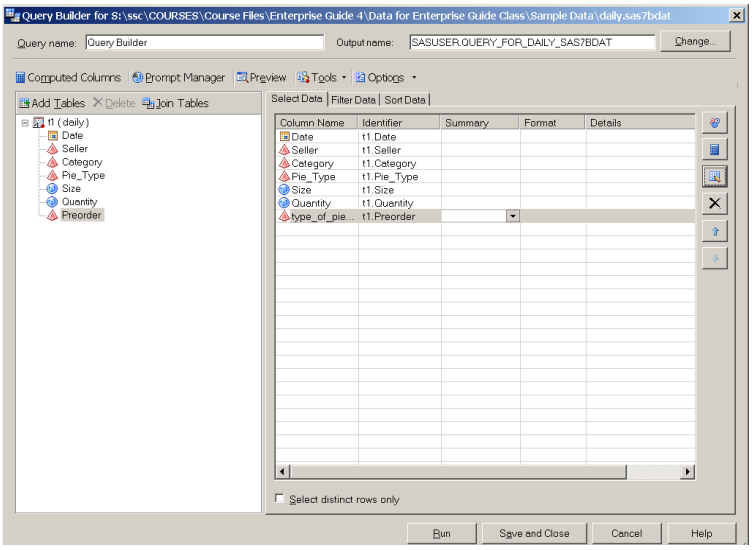


Saving, Running and Canceling

Saving stores the query so it can be modified or run later.

Running carries out the query and also saves it.

Canceling closes the query without saving it.



Query Results Format

There are three formats for query results:

Data table

Data view

Report (HTML)

Data Tables

A Data Table:

Is a static table of values.

Is not updated unless the query is rerun.

Can have tasks run against it.

	Date	Seller	Category	Pie_Type	Size	Quantity	Preorder
1	06/25/2006	Betty	Chocolate Pies	Death By Chocola	8	1	No
2	06/25/2006	Betty	Fruit Pies	Lemon Meringue	10	5	Yes
3	06/25/2006	Lenore	Fruit Pies	Strawberry	10	2	No
4	06/25/2006	Frank	Fruit Pies	Apple	10	1	No
5	06/25/2006	Betty	Fruit Pies	Blueberry	10	3	Yes
6	06/25/2006	Marge	Chocolate Pies	French Silk	8	2	No
7	06/25/2006	Marge	Chocolate Pies	Peanut Butter Cup	10	1	No
8	06/25/2006	Frank	Chocolate Pies	Black Forest	8	5	Yes
9	06/25/2006	Lenore	Fruit Pies	Banana Cream	8	2	Yes
10	06/25/2006	Betty	Chocolate Pies	Peanut Butter Cup	8	2	No
11	06/25/2006	Frank	Fruit Pies	Raspberry	10	2	No
12	06/25/2006	Marge	Fruit Pies	Apple	8	4	Yes

Data Views

A Data View:

Is a dynamic table.

Stores logic to carry out on data rather than the data itself.

Is updated as the source data changes.

Can have tasks run against it.

Reports

A Report:

Is a static HTML document.

Is not updated unless the query is rerun.

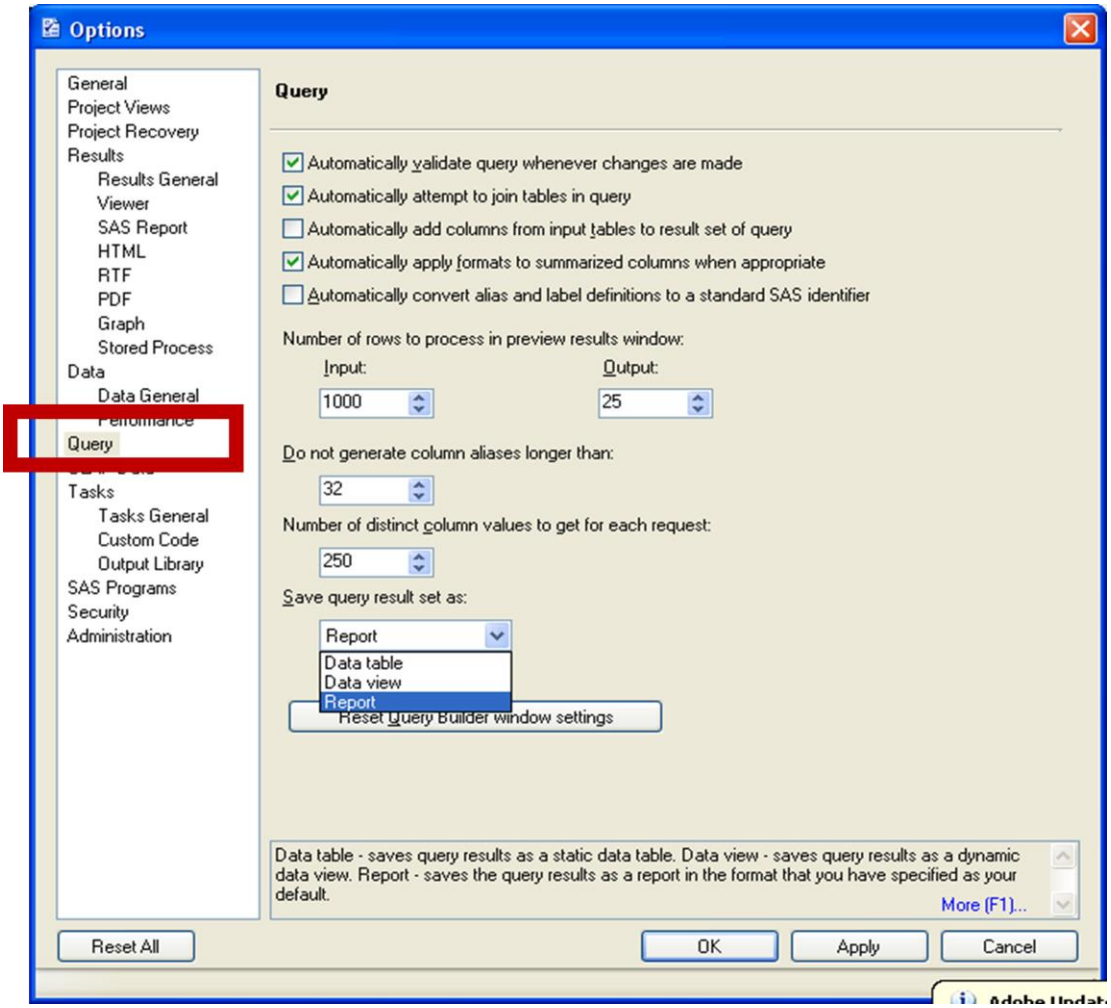
Can not have have tasks run against it.

Date	Seller	Category	Pie_Type	Size	Quantity	Preorder
06/25/2006	Betty	Chocolate Pies	Death By Chocolate	8	1	No
06/25/2006	Betty	Fruit Pies	Lemon Meringue	10	5	Yes
06/25/2006	Lenore	Fruit Pies	Strawberry	10	2	No
06/25/2006	Frank	Fruit Pies	Apple	10	1	No
06/25/2006	Betty	Fruit Pies	Blueberry	10	3	Yes
06/25/2006	Marge	Chocolate Pies	French Silk	8	2	No
06/25/2006	Marge	Chocolate Pies	Peanut Butter Cup	10	1	No
06/25/2006	Frank	Chocolate Pies	Black Forest	8	5	Yes
06/25/2006	Lenore	Fruit Pies	Banana Cream	8	2	Yes
06/25/2006	Betty	Chocolate Pies	Peanut Butter Cup	8	2	No
06/25/2006	Frank	Fruit Pies	Raspberry	10	2	No
06/25/2006	Marge	Fruit Pies	Apple	8	4	Yes

Selecting Query Results Format

On the Tools menu, select Options, Query.

Save query result set as, Select Data Table, Data View or Report.

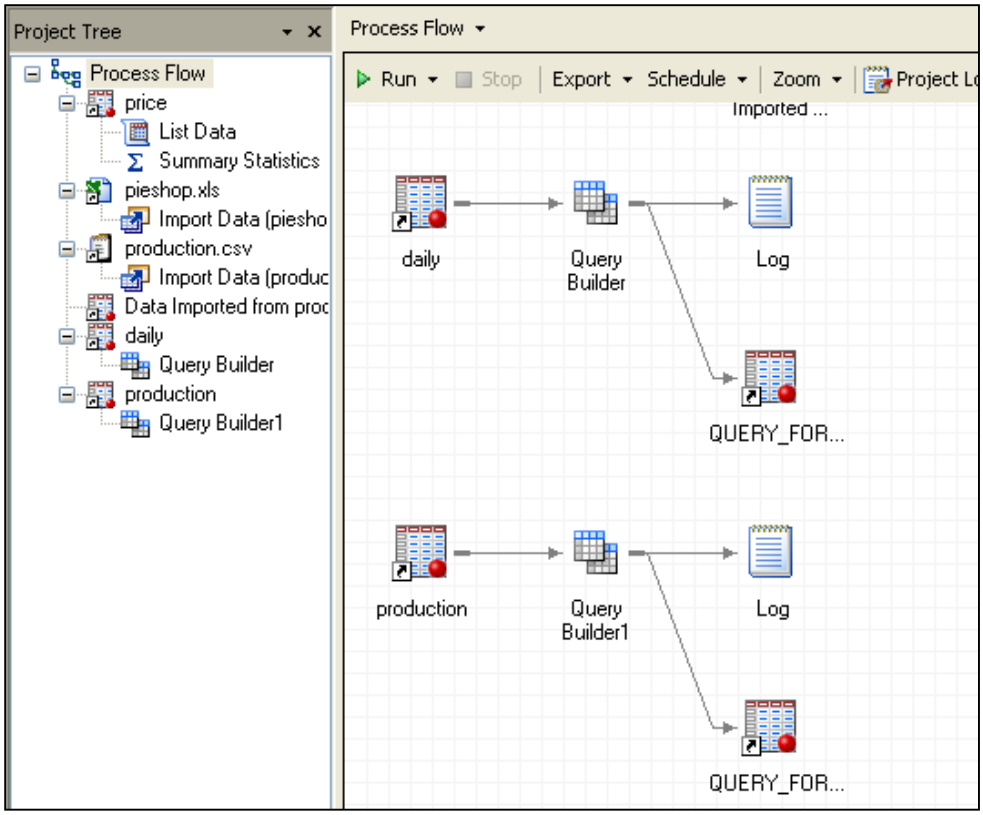


Query Results

After the query runs:

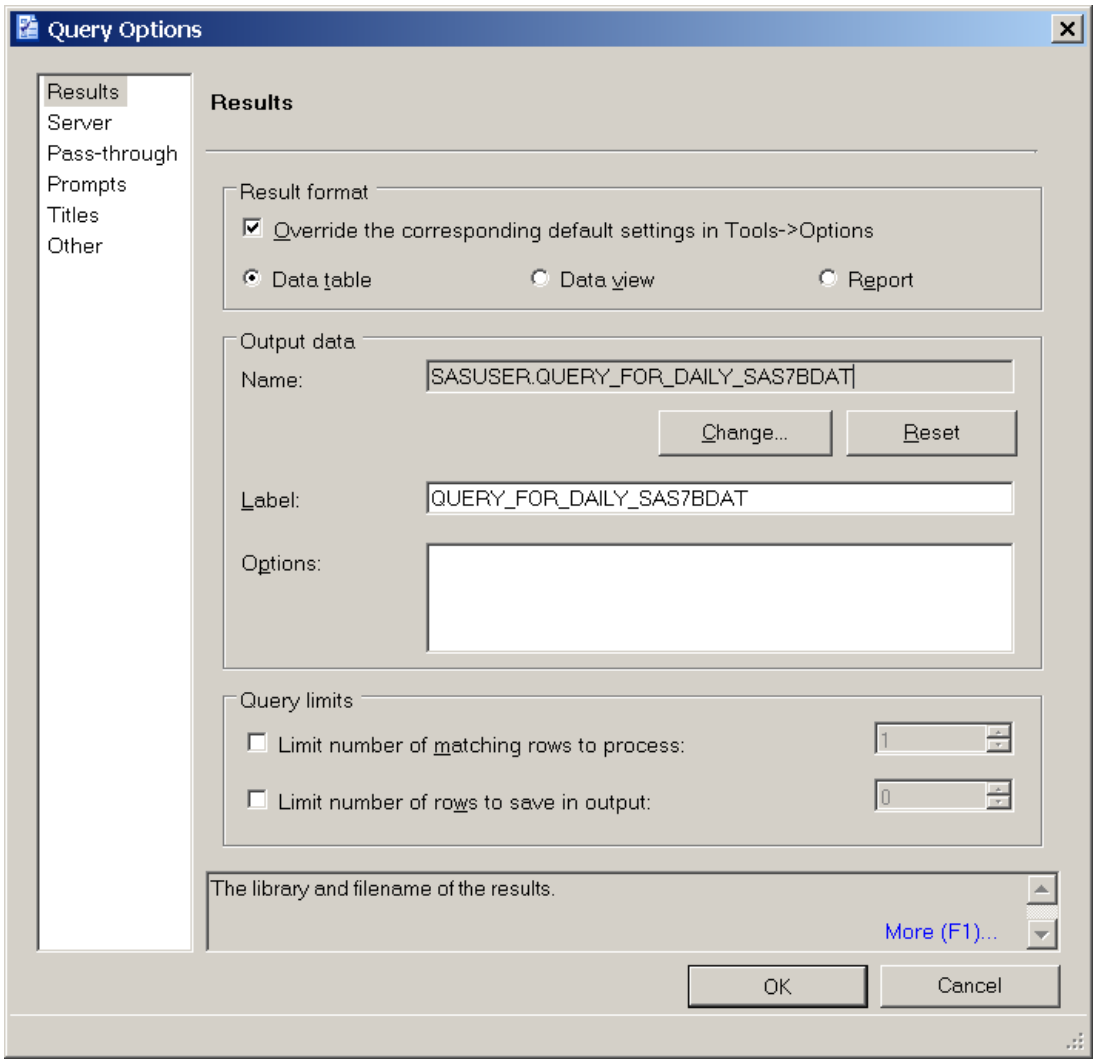
Objects for the query and its results appear in the Process Flow and Project Tree.

The original data set has not changed.



Working with Query Results

You can set the output location of your query, under the options.



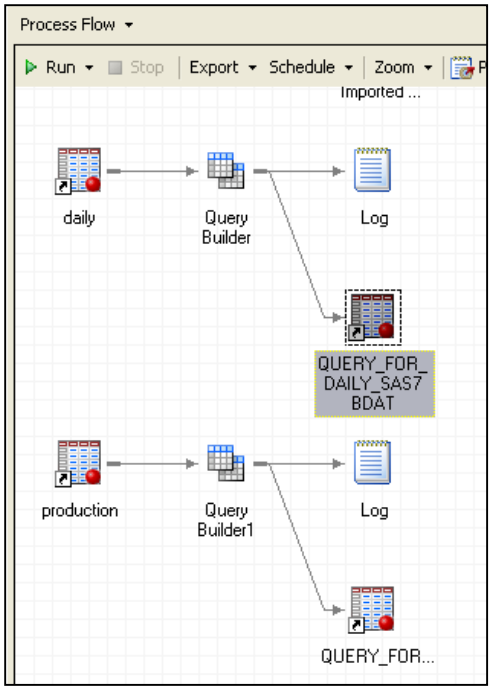
Working with Query Results

Tasks may be run on query results if you have specified data table or data view as your output option.

To run a task on query results:

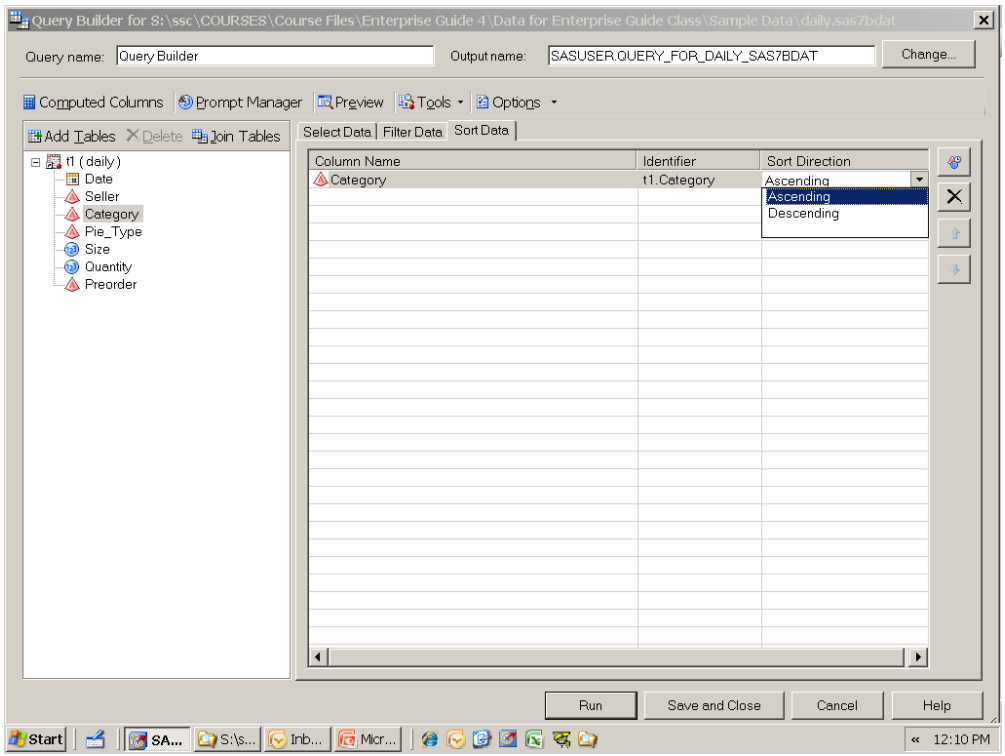
In the Process Flow or Project Tree, highlight the query results.

From the Task List, select a task and proceed as usual.



Sorting

Sorting changes the order of the observations in a table by ranking the observations based on the values of one or more variables.



Results

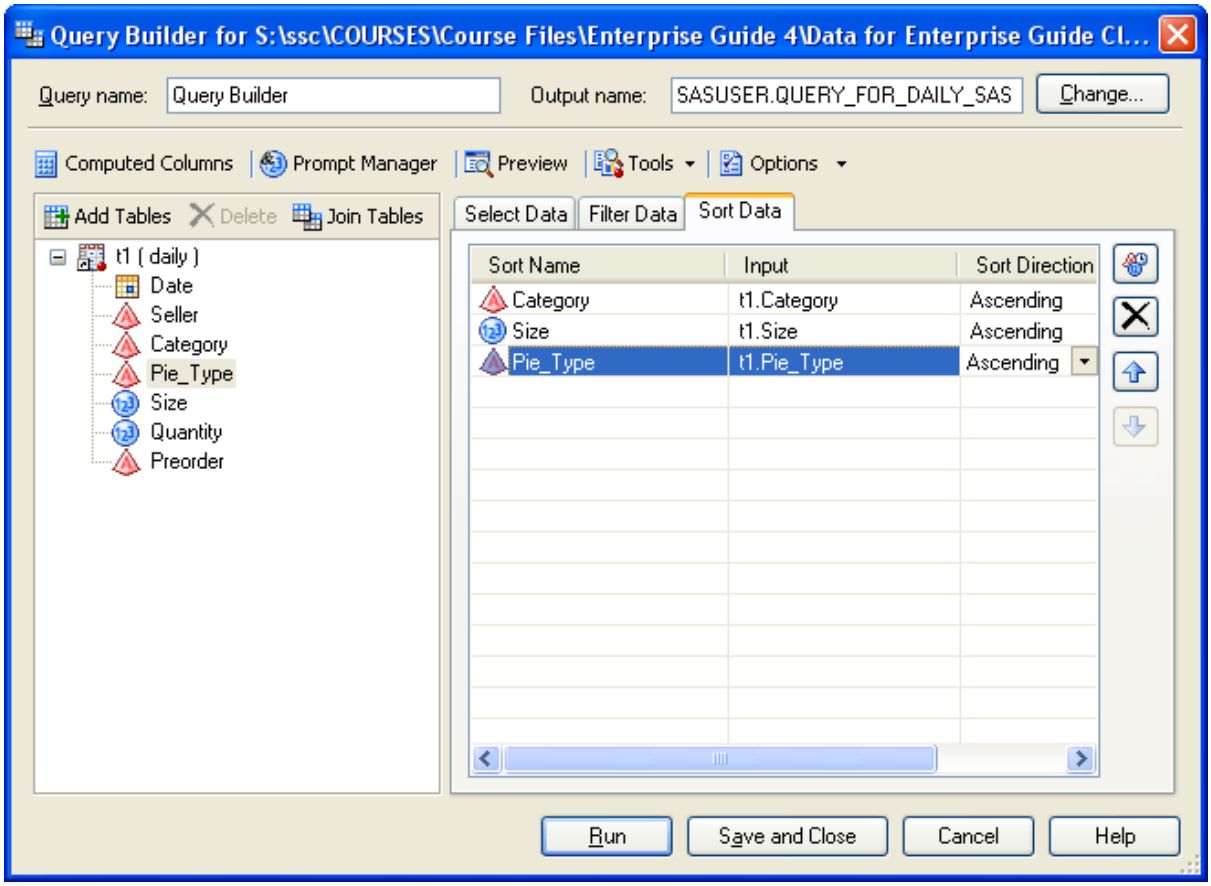
Click Run to run the query.

Category	Pie Type	Size	Quantity	Seller
Chocolate Pies	Death By Chocolate	8	1	Betty
Chocolate Pies	Peanut Butter Cup	8	2	Betty
Chocolate Pies	Peanut Butter Cup	10	1	Marge
Chocolate Pies	French Silk	8	2	Marge
Chocolate Pies	Black Forest	8	5	Frank
Fruit Pies	Strawberry	10	2	Lenore
Fruit Pies	Apple	8	4	Marge
Fruit Pies	Blueberry	10	3	Betty
Fruit Pies	Raspberry	10	2	Frank
Fruit Pies	Lemon Meringue	10	5	Betty
Fruit Pies	Apple	10	1	Frank
Fruit Pies	Banana Cream	8	2	Lenore

Sorting by Multiple Variables

Add additional variables to the Sort Data tab.

Use the  and  buttons to arrange the variables in order of sort priority.



Results

Use the Select Data tab to change the order of the variables.

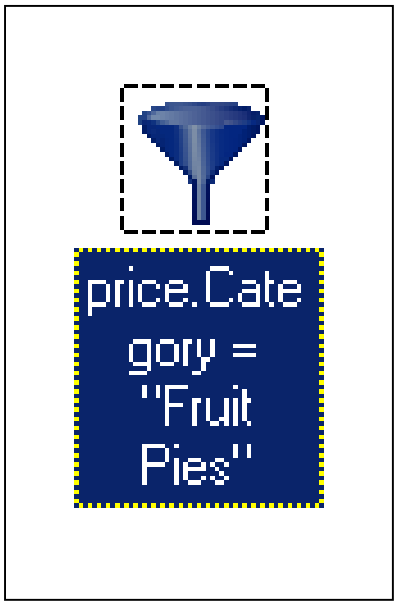
Click Run to run the query.

Category	Size	Pie_Type	Quantity	Seller
Chocolate Pies	8	Black Forest	5	Frank
Chocolate Pies	8	Death By Chocolate	1	Betty
Chocolate Pies	8	French Silk	2	Marge
Chocolate Pies	8	Peanut Butter Cup	2	Betty
Chocolate Pies	10	Peanut Butter Cup	1	Marge
Fruit Pies	8	Apple	4	Marge
Fruit Pies	8	Banana Cream	2	Lenore
Fruit Pies	10	Apple	1	Frank
Fruit Pies	10	Blueberry	3	Betty
Fruit Pies	10	Lemon Meringue	5	Betty
Fruit Pies	10	Raspberry	2	Frank
Fruit Pies	10	Strawberry	2	Lenore

Filtering Data

Filtering is carried out in the Query Builder.

Note that there is also a task for just filtering and sorting.



What is a Filter?

A filter:

Specify a condition that determines which rows will be included in the query results.

Does not affect which columns are in the results.

Examples:

Inventory is less than 200

Age is greater than 65

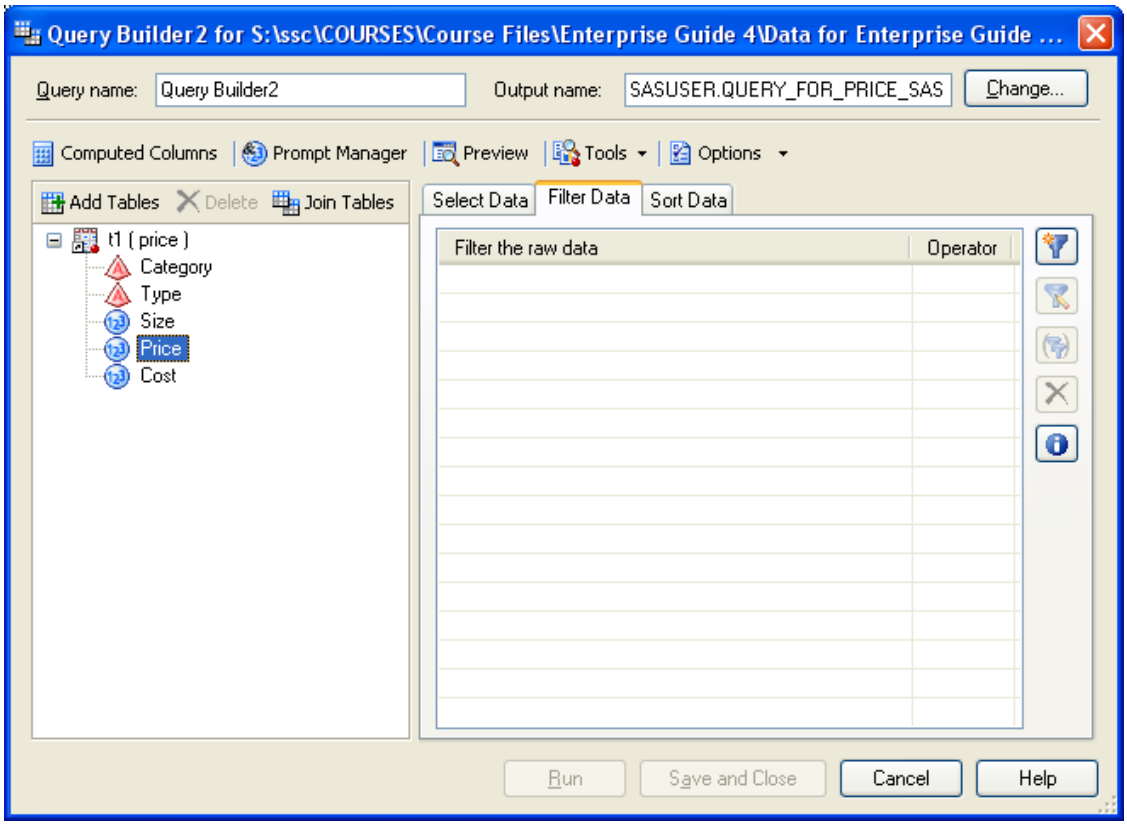
Importance equals ‘urgent’.

Creating a New Filter in a Query

Click the Filter Data tab.

Double-click a variable to open the Edit Filter window.

This variable will be the left side of the filter condition.

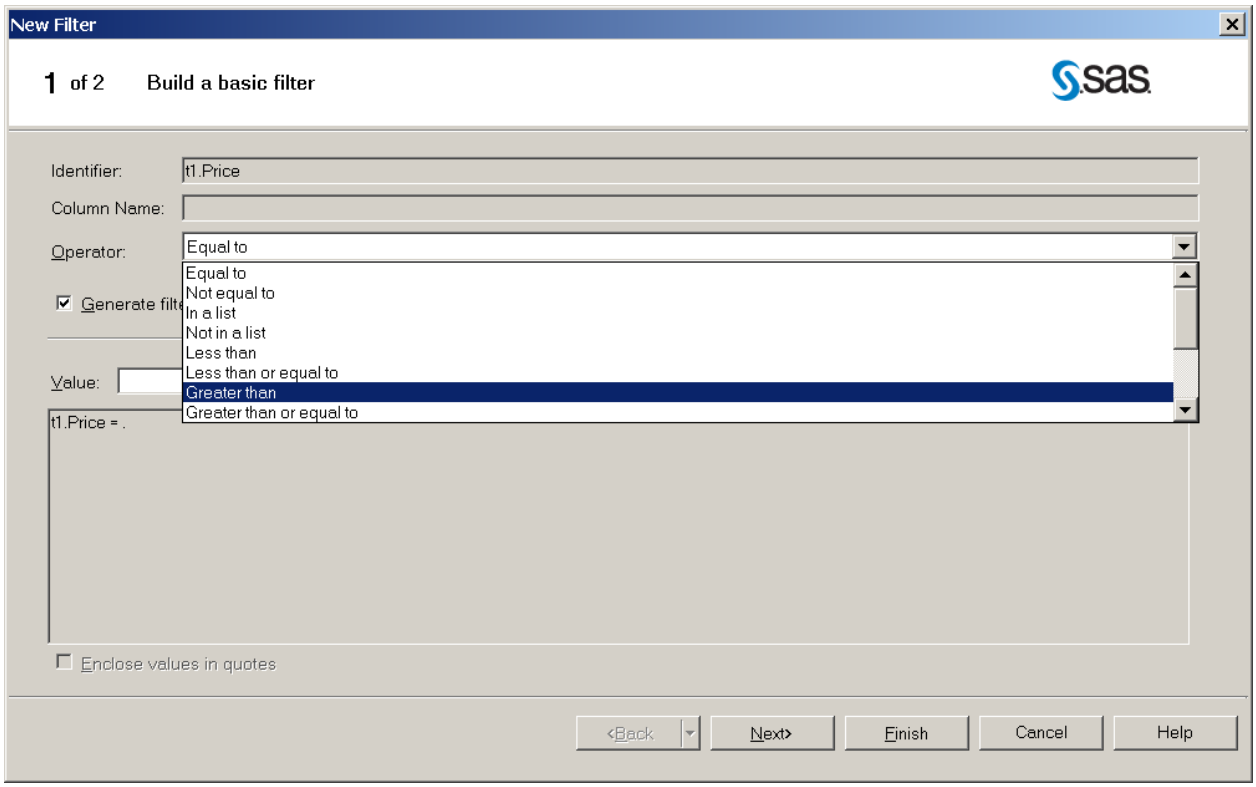


Operators

An operator states the relationship between the two sides of the filter condition.

Examples of operators are greater than, less than, and equal to.

Selecting an Operator



Select an operator from the Operator drop-down menu.

The Right Side of the Filter Condition

The right side of the filter condition may be:

A value

A value selected from a list

A group of values

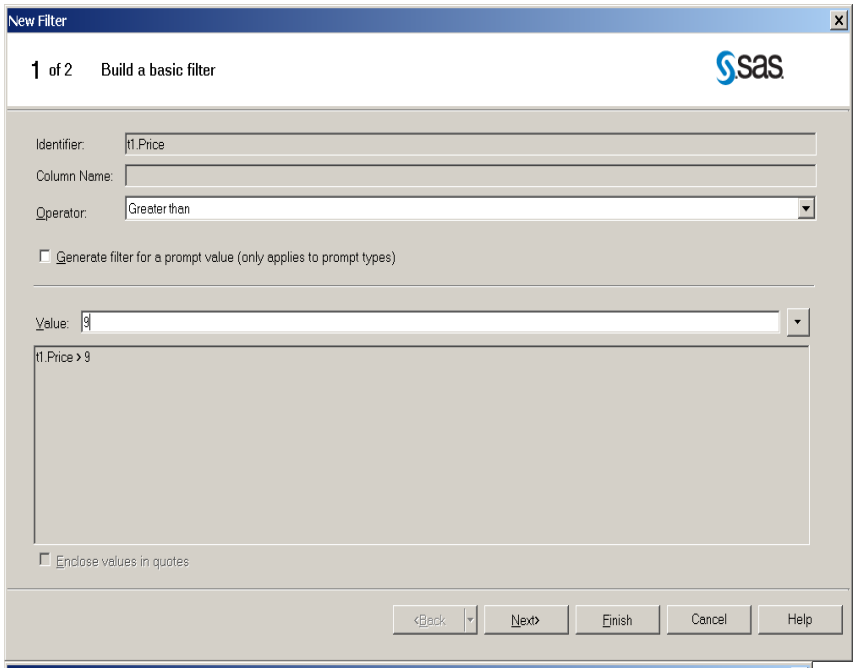
A variable

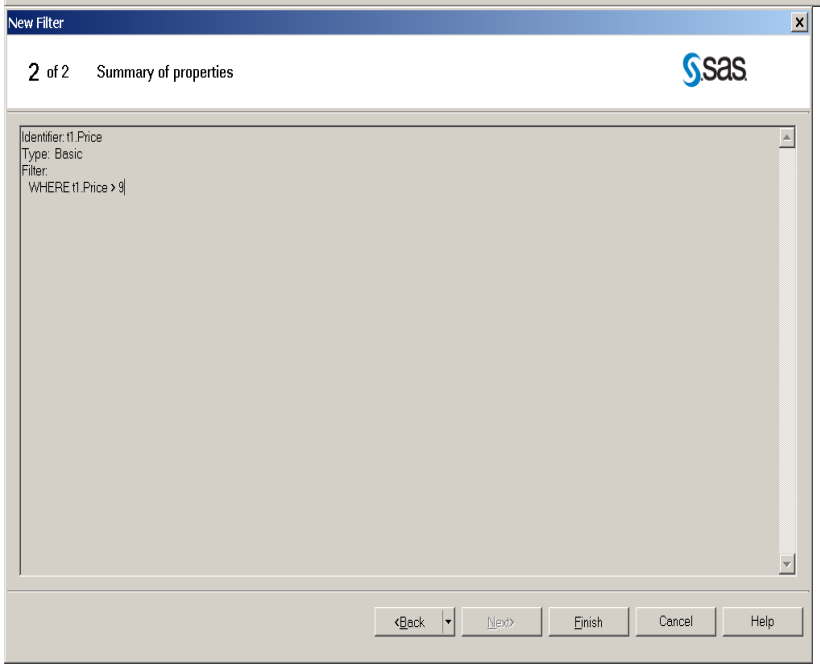
A prompt

Right Side: User-Entered Value

Type a value in the Value field.

Click Next.





Verify Summary of properties and click Finish.

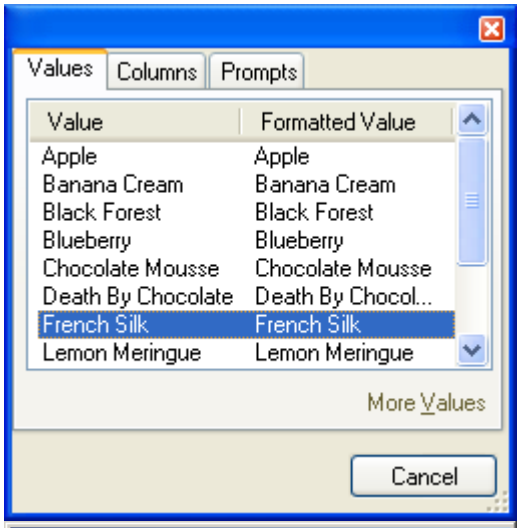
Right Side: Value from a List

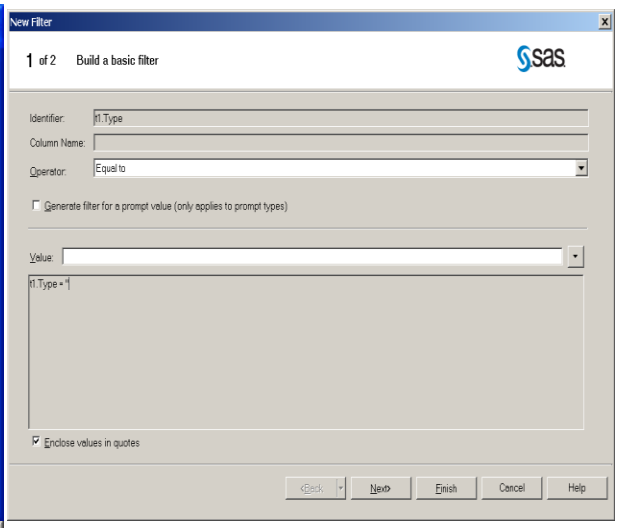
In the New Filter window, click the  button next to the Value field.

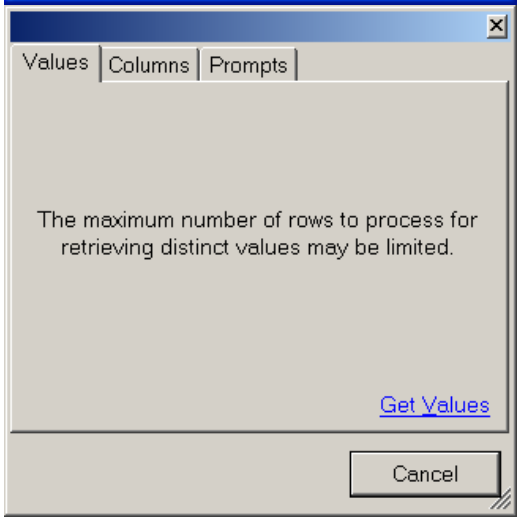
On the Values tab, click Get Values.

Select a value on the list.

Verify Summary of properties and click Finish.





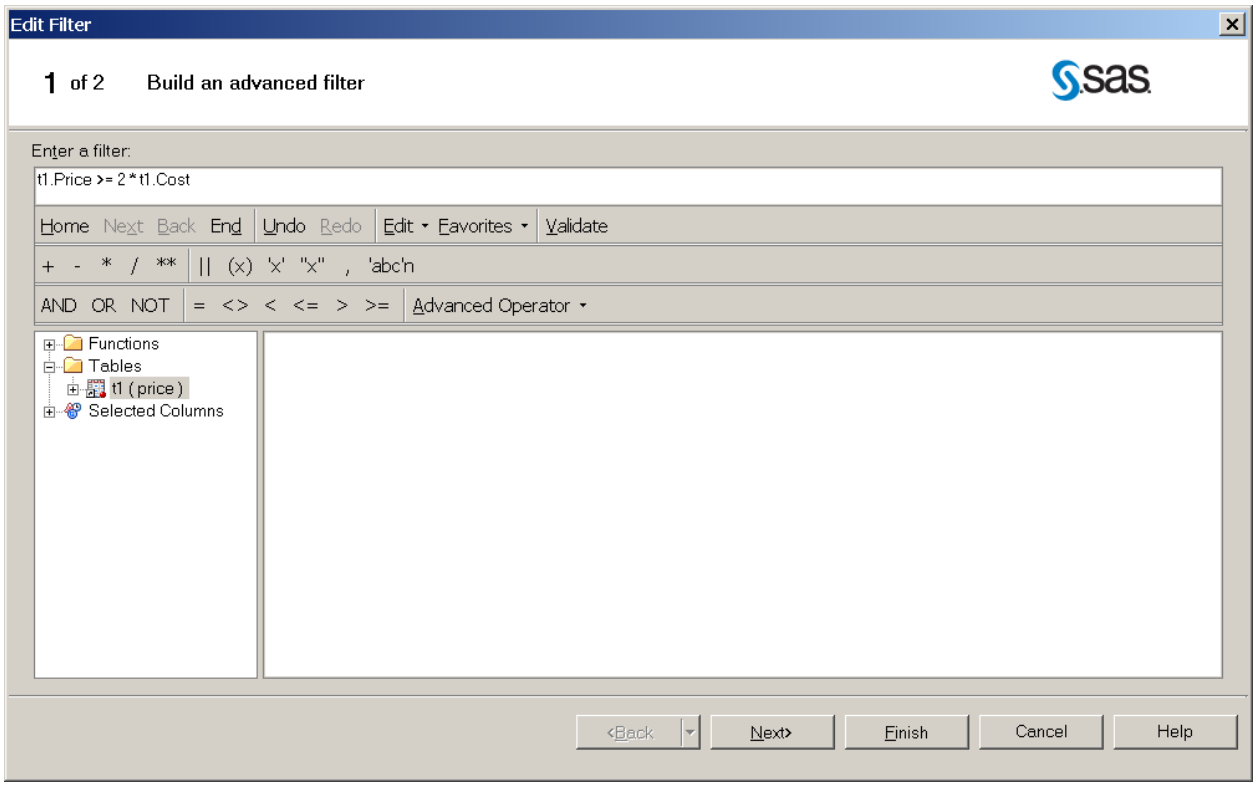


Advanced Filters

Advanced filters can contain expressions or functions on either side of the operator.

Advanced filters are created using the Advanced Expression Builder.

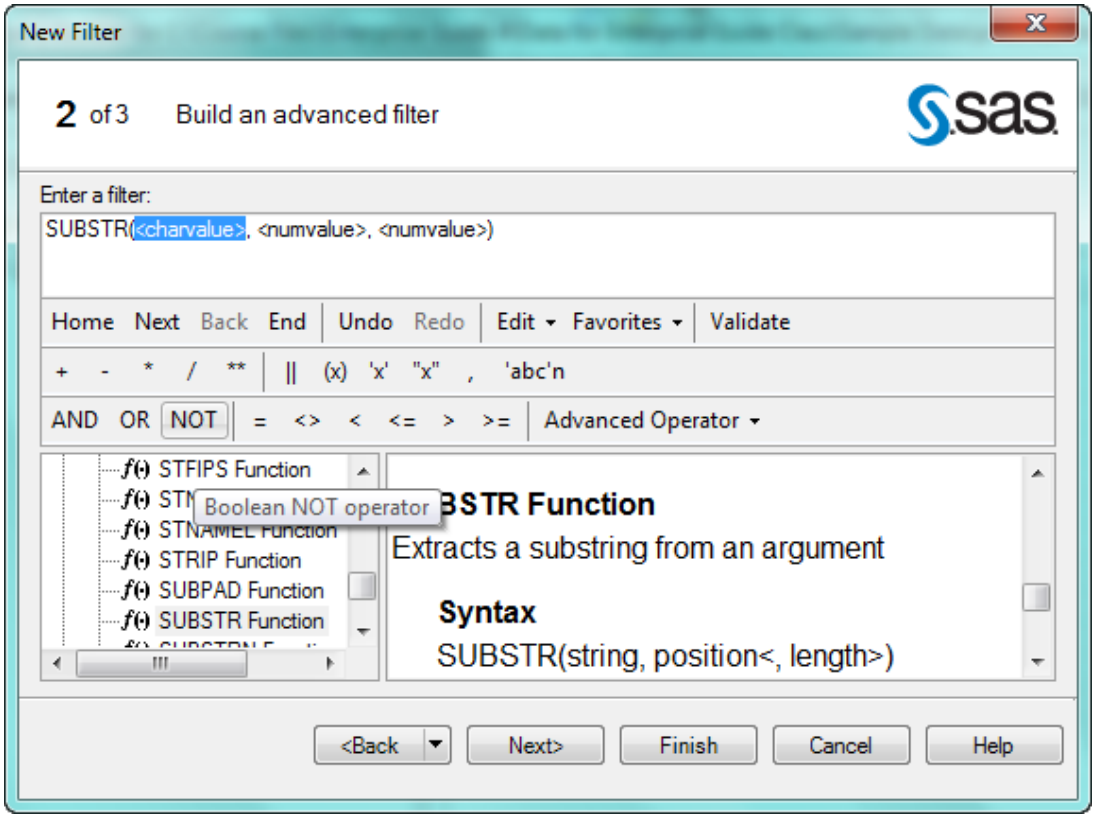
Build by typing in the expression box, double-clicking variables, values, or functions, or single-clicking operators.



Functions In A Filter

Over 500 SAS functions are available in a filter.

Scroll down to SUBSTR, click. The first argument can be typed over, or double clicked and a column chosen.



Multiple Filters: AND/OR

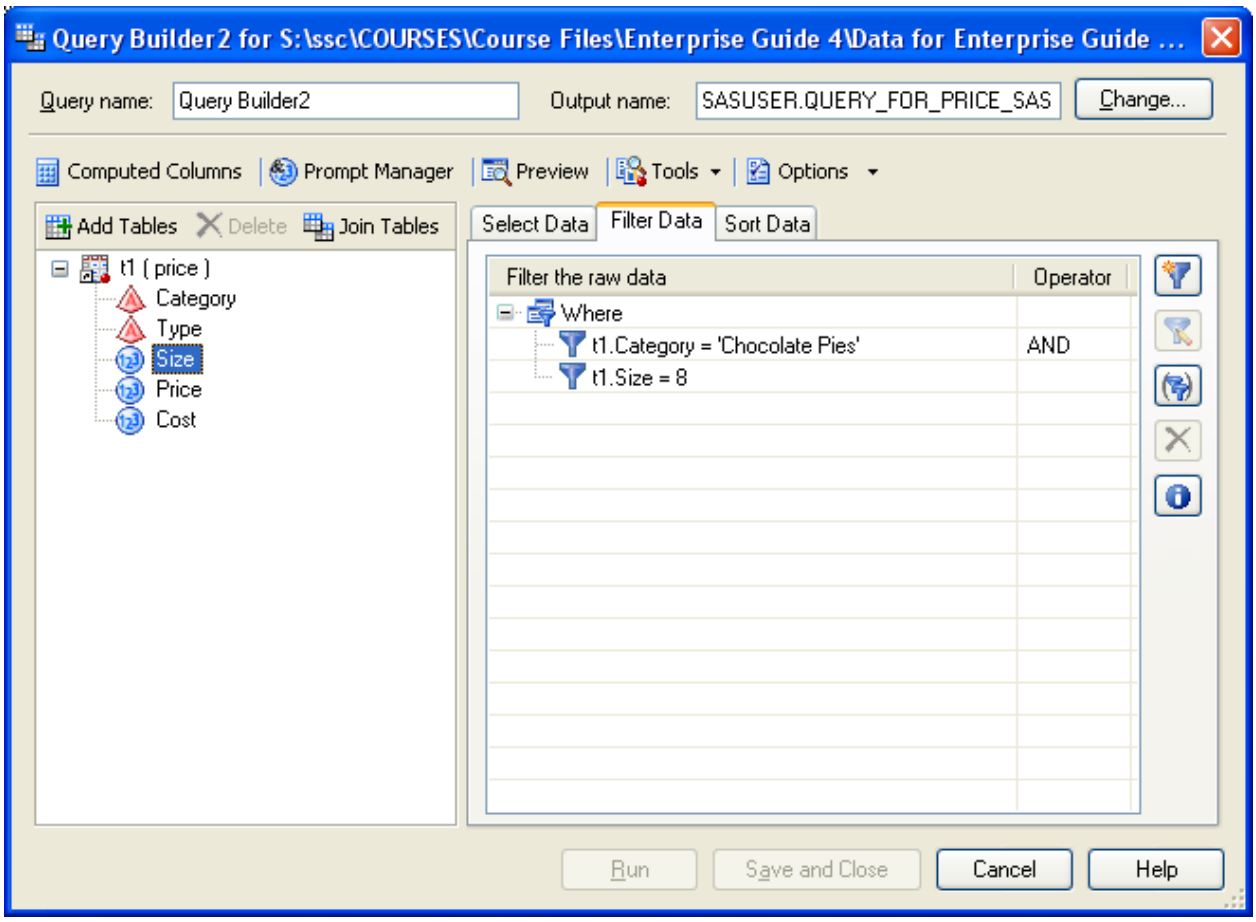
Ways filters can relate to each other:

AND: An observation will pass through only if both conditions are true.

OR: An observation will pass through if either condition is true or if both are true.

AND Filters

By default, if two or more filters are added to a query, the filters are related by AND.



Results

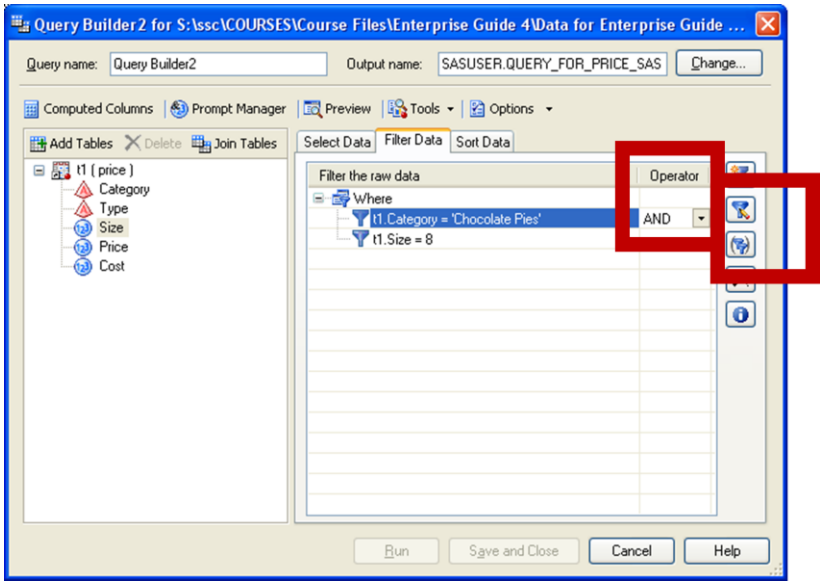
Click Run to run the query.

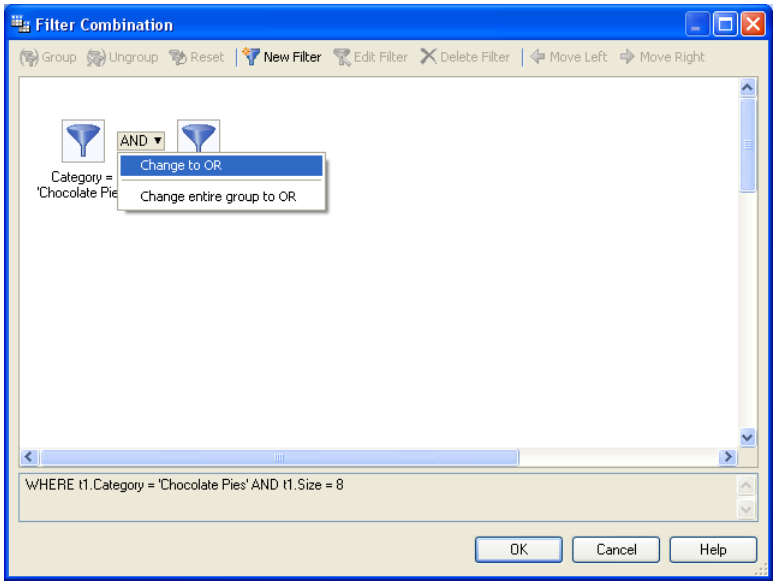
Category	Type	Size	Price	Cost
Chocolate Pies	Chocolate Mousse	8	8.99	4.25
Chocolate Pies	French Silk	8	6.99	3.2
Chocolate Pies	Death By Chocolate	8	10.99	5.25
Chocolate Pies	Peanut Butter Cup	8	9.99	5
Chocolate Pies	Black Forest	8	7.99	3.95

Changing an AND Filter to an OR Filters

On the Filter Data tab, click Combine Filters  to open the Filter Combination window.

Click AND between the two filters. Click Change to OR.





You can also click on the down arrow next to the And to modify.

Results

Click Run to run the query.

Category	Type	Size	Price	Cost
Chocolate Pies	Chocolate Mousse	8	8.99	4.25
Chocolate Pies	Chocolate Mousse	10	10.99	5.5
Chocolate Pies	French Silk	8	6.99	3.2
Chocolate Pies	French Silk	10	8.99	4.75
Chocolate Pies	Death By Chocolate	8	10.99	5.25
Chocolate Pies	Death By Chocolate	10	12.99	6.85
Chocolate Pies	Peanut Butter Cup	8	9.99	5
Chocolate Pies	Peanut Butter Cup	10	11.99	6
Chocolate Pies	Black Forest	8	7.99	3.95
Chocolate Pies	Black Forest	10	9.99	4.95
Fruit Pies	Raspberry	8	6.99	3.55
Fruit Pies	Rhubarb	8	5.99	3.15
Fruit Pies	Strawberry	8	4.99	2.6
Fruit Pies	Blueberry	8	5.99	3.2
Fruit Pies	Lemon Meringue	8	8.99	4.55
Fruit Pies	Banana Cream	8	7.99	4.2
Fruit Pies	Apple	8	6.99	3.7

Additional Filter Hints

Grouping Filters

Note the difference between:

(Filter 1 AND Filter 2) OR Filter 3

and

Filter 1 AND (Filter 2 OR Filter 3)

In the Filter Combination window, hold down the CTRL key and click on filters to select filters for grouping.

Use the Group and Ungroup buttons to change the way the selected filters are grouped.

Deleting Filters

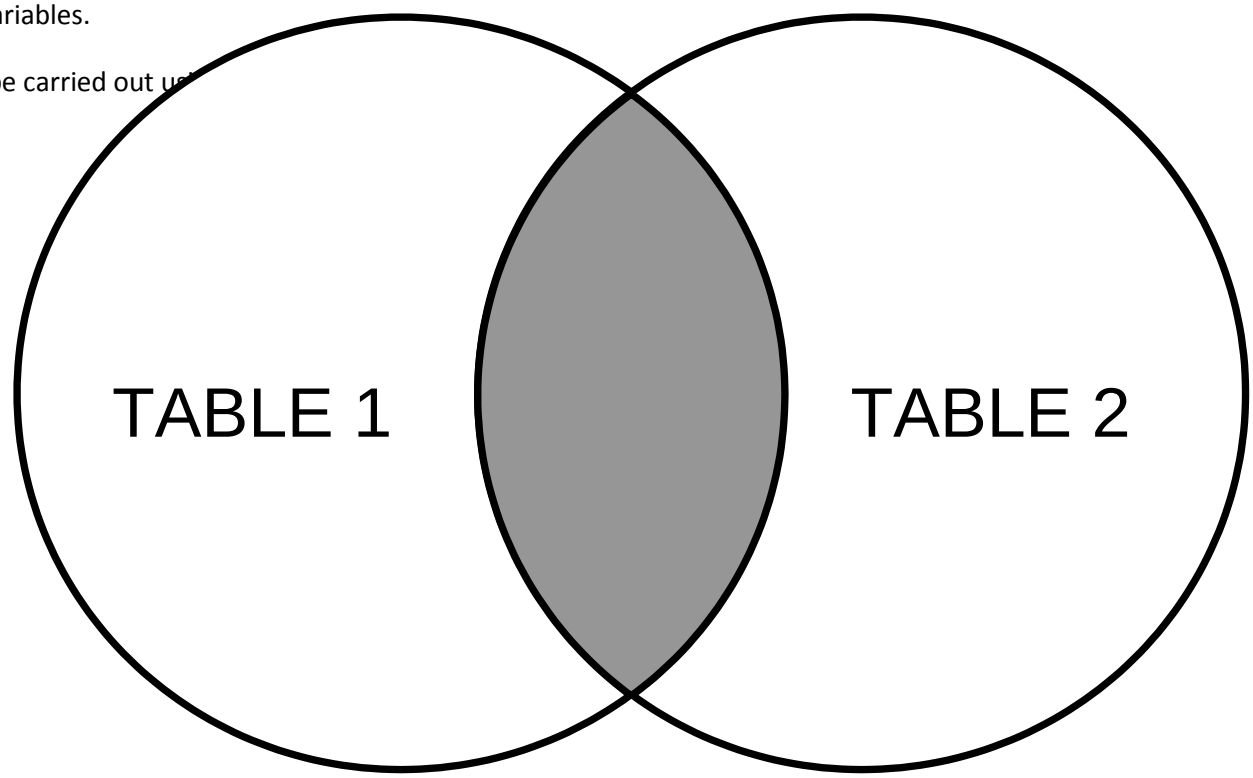
On the Filter Data tab, highlight a filter.

Click the X button to delete the filter.

Joining Data Tables

Joining is the process of merging two data sets by fusing together certain observations based on the values of variables.

A join can be carried out u



Automatic Joins

An automatic join occurs when two data sets in a query have a variable with the same name and type.

EG searches for a suitable join variable for an automatic join when the second data table is added to the query.

An automatic join is an equijoin by default.

Inner Joins

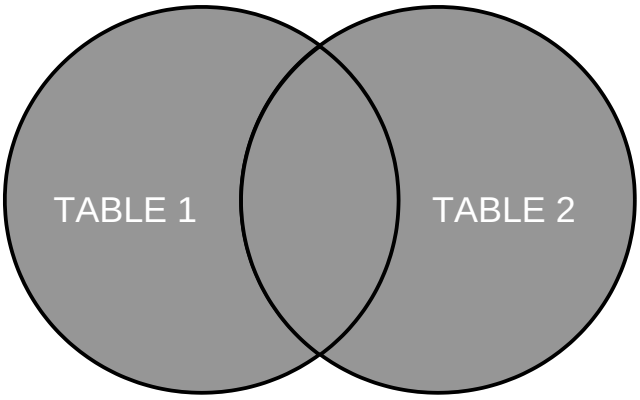
In an inner join, a row is included only if the join variable value is common to both tables.

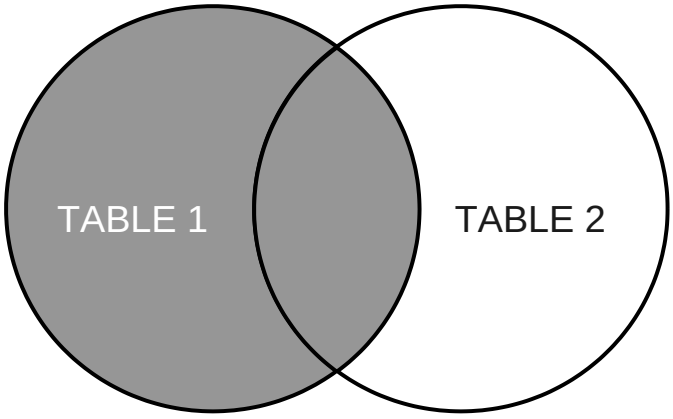
Rows without a match are omitted.

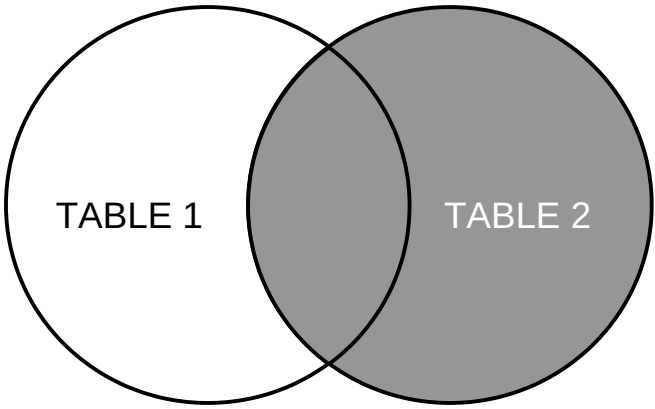
Outer Joins

In a full outer join, all rows from both tables are included.

In a right outer join or a left outer join, all rows from one of the tables are included, along with the corresponding rows from the other table.







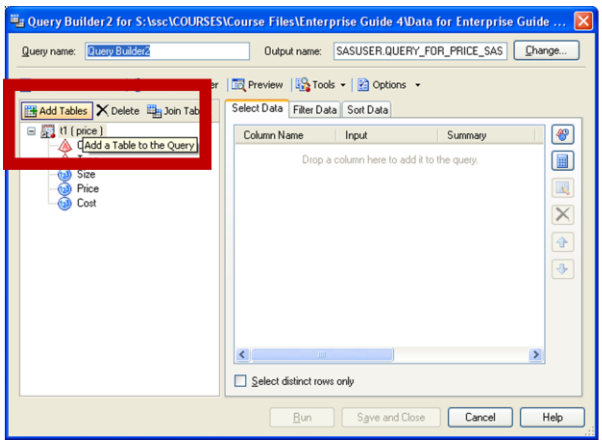
Adding an Additional Data Set to a Query

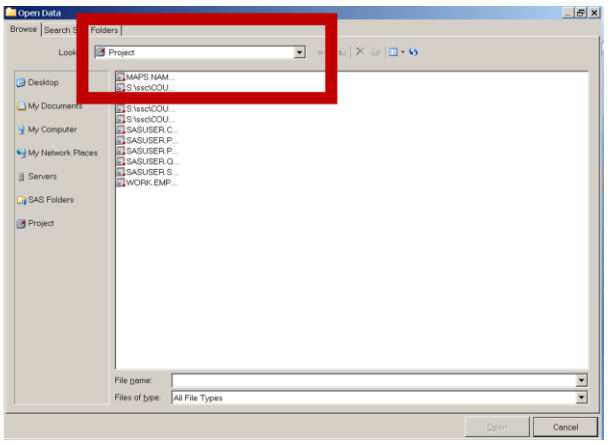
Click Add Tables.

Click Local Computer, Servers, SAS Folders, or Project.

Highlight the name of the data set.

Click Open.

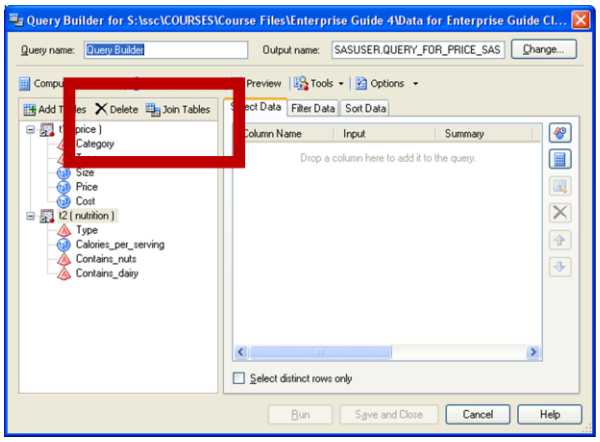


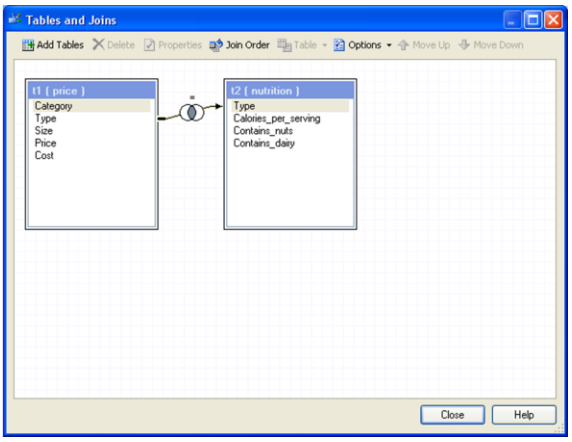


Viewing Automatic Joins

Click the Join Tables button.

The diagram will indicate which columns will be used as join variables.

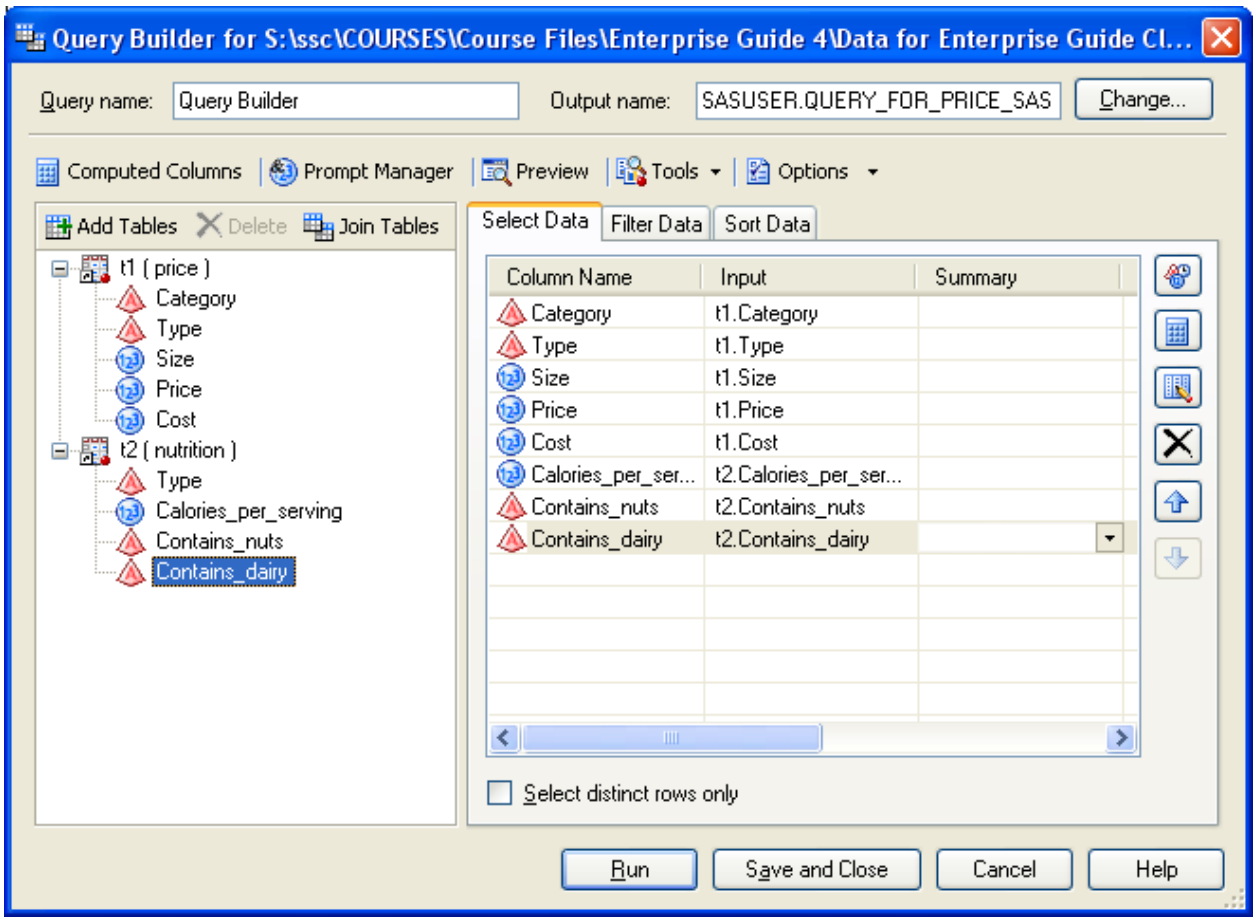




Selecting Columns

By default, none of the columns from the new table will be included in the query.

Double-click columns on the Tables list to add them to the query.



Results

Click Run to run the query.

Category	Type	Size	Price	Calories per Serving	Contains Nuts	Contains Dairy
Chocolate Pies	French Silk	8	8.99	550	No	Yes
Chocolate Pies	French Silk	10	10.99	550	No	Yes
Chocolate Pies	French Silk	8	6.99	475	No	Yes
Chocolate Pies	French Silk	10	8.99	475	No	Yes
Chocolate Pies	Death By Chocolate	8	10.99	800	No	Yes
Chocolate Pies	Death By Chocolate	10	12.99	800	No	Yes
Chocolate Pies	Peanut Butter Cup	8	9.99	750	Yes	Yes
Chocolate Pies	Peanut Butter Cup	10	11.99	750	Yes	Yes
Chocolate Pies	Black Forest	8	7.99	700	No	Yes
Chocolate Pies	Black Forest	10	9.99	700	No	Yes
Fruit Pies	Raspberry	8	6.99	400	No	No
Fruit Pies	Raspberry	10	8.99	400	No	No
Fruit Pies	Rhubarb	8	5.99	425	No	No
Fruit Pies	Rhubarb	10	7.99	425	No	No
Fruit Pies	Strawberry	8	4.99	485	No	No
Fruit Pies	Strawberry	10	6.99	485	No	No
Fruit Pies	Blueberry	8	5.99	375	No	No
Fruit Pies	Blueberry	10	7.99	375	No	No
Fruit Pies	Lemon Meringue	8	8.99	525	No	Yes
Fruit Pies	Lemon Meringue	10	10.99	525	No	Yes
Fruit Pies	Banana Cream	8	7.99	490	No	Yes
Fruit Pies	Banana Cream	10	9.99	490	No	Yes
Fruit Pies	Apple	8	6.99	350	No	No
Fruit Pies	Apple	10	8.99	350	No	No

Manual Joins

A manual join:

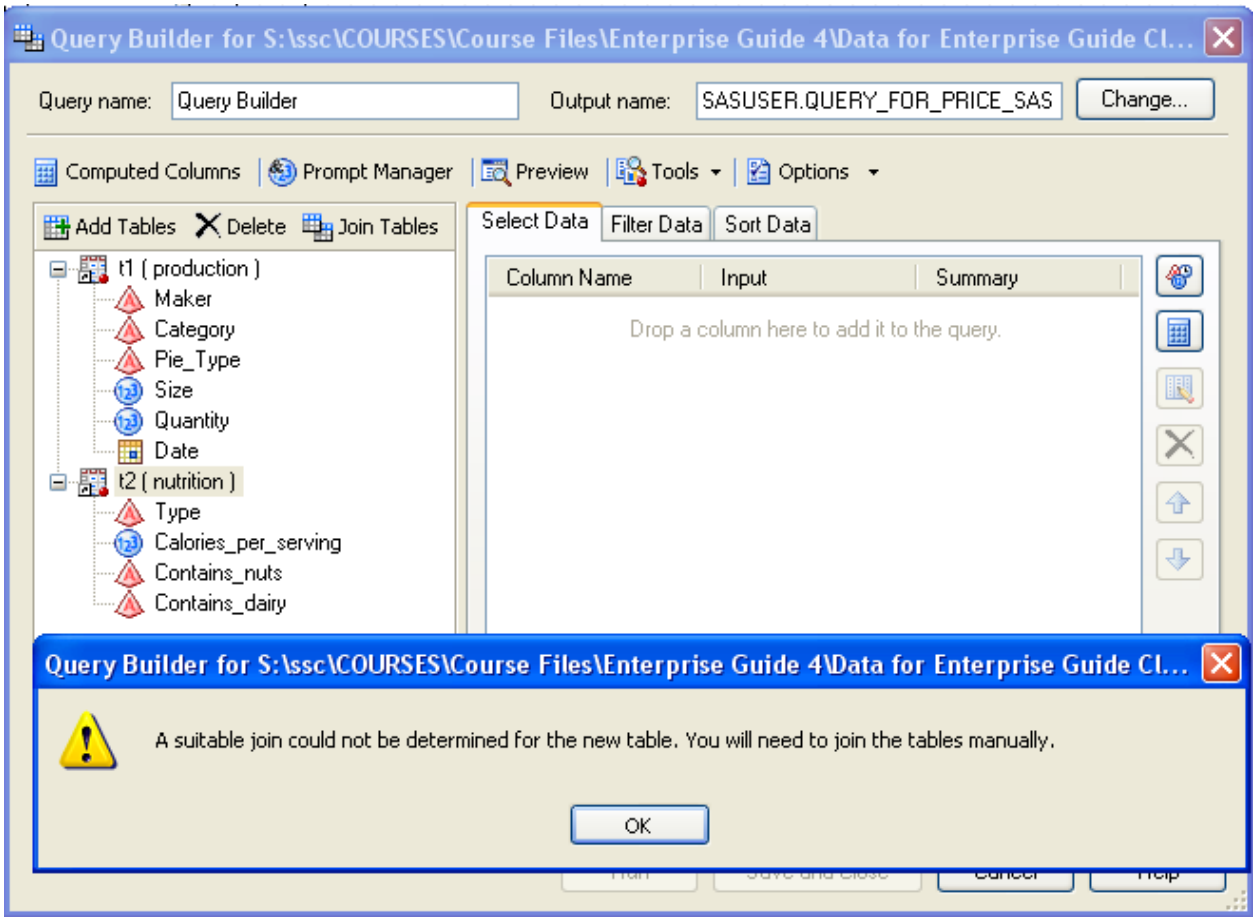
Requires you to specify which variables will be used to join the tables.

Is necessary when joining by more than one variable, or variables with different names.

Performing a Manual Join

Add the second data set.

Click OK when a message tells you that you need to join the tables manually.



Performing a Manual Join (continued)

Right-click the join variable from the first table.

Select Join <variable> with.

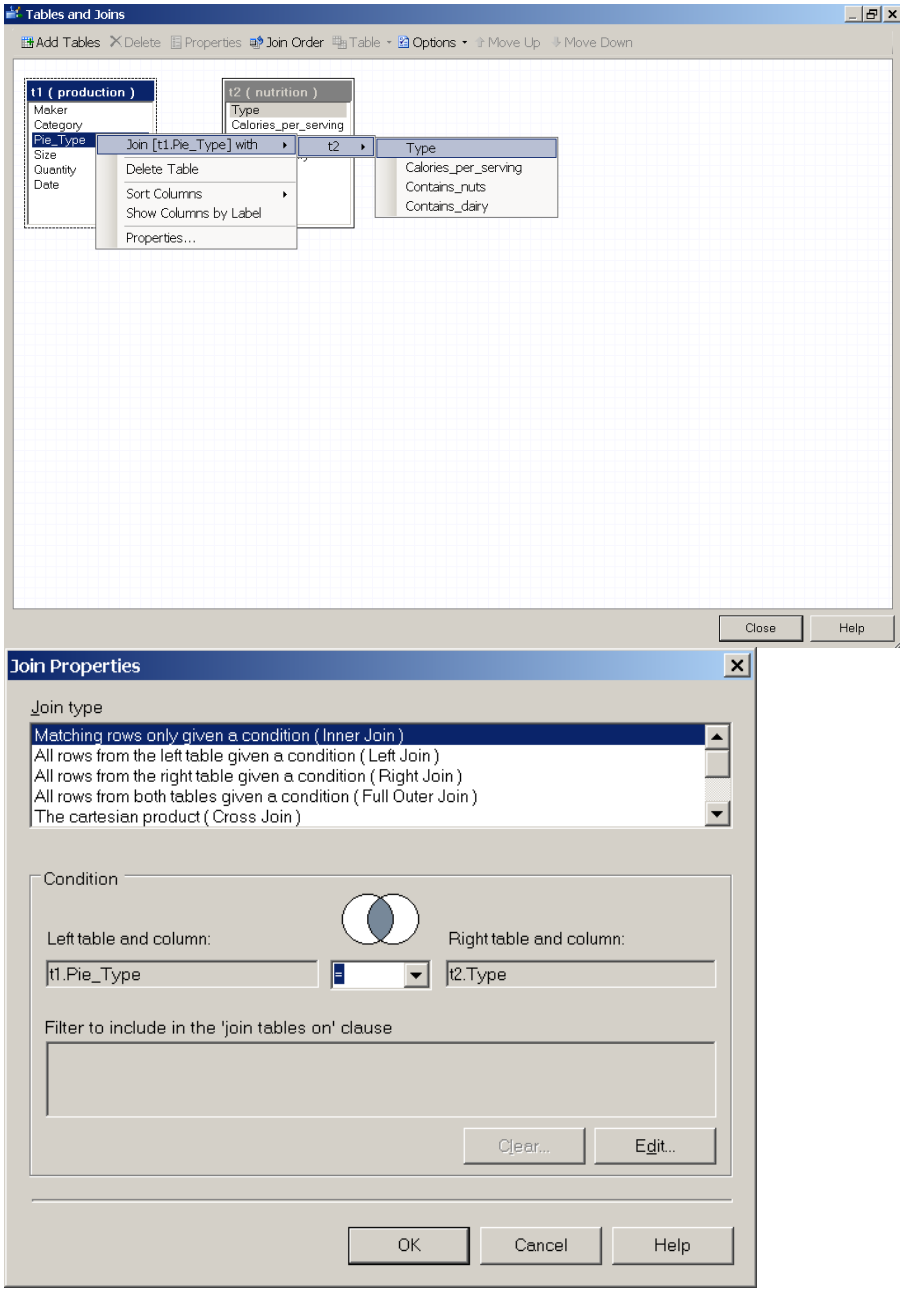
Select the name of the second table.

Select the name of the join variable from the second table.

A window will pop up for you to define what type of join.

Click OK to close the Join Properties screen.

Click CLOSE to close the Tables and Join screen.



Multiple Common Variables

If data sets have multiple variables in common:

An automatic join uses only one variable to match the data sets.

It may be necessary to manually join the tables using other variables the data sets have in common.

Joining by Multiple Variables (continued)

Right-click the next join variable from the first table.

Select Join <variable> with.

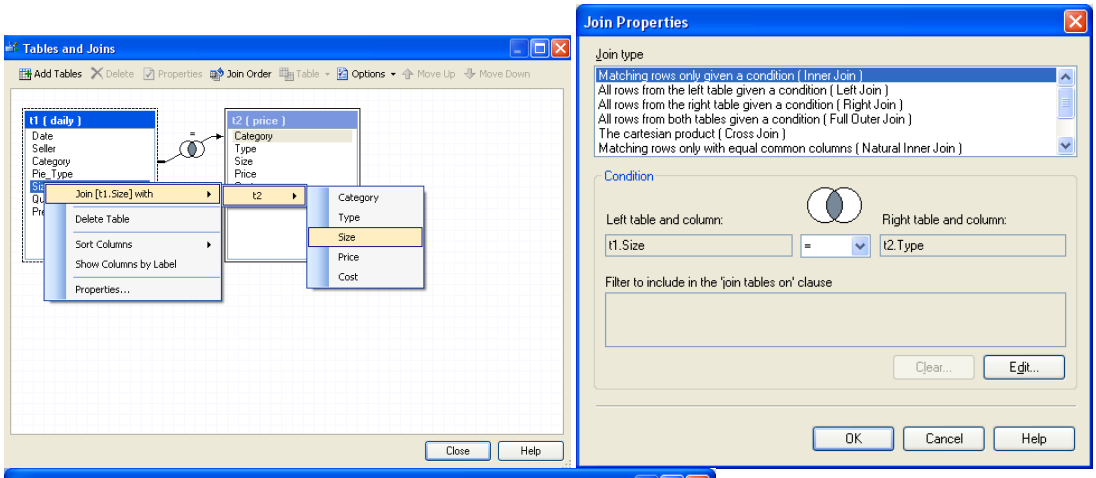
Select the name of the second table.

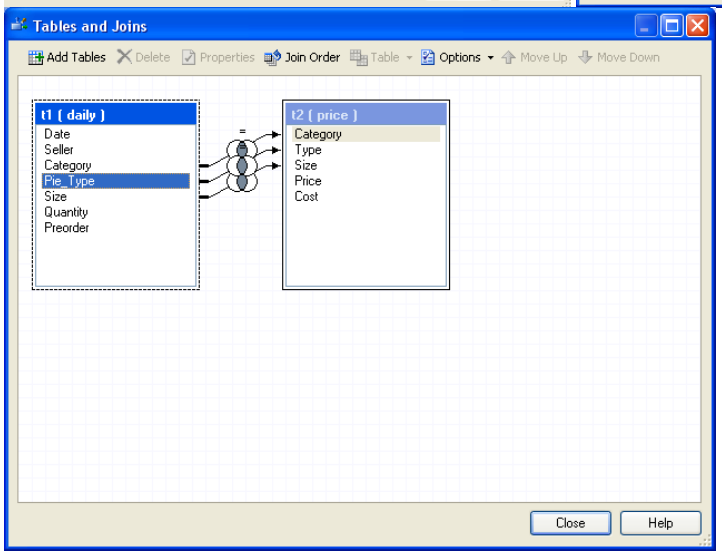
Select the name of the join variable from the second table.

Select the type of join from the pop-up menu.

Repeat for additional join variables.

Click Close.



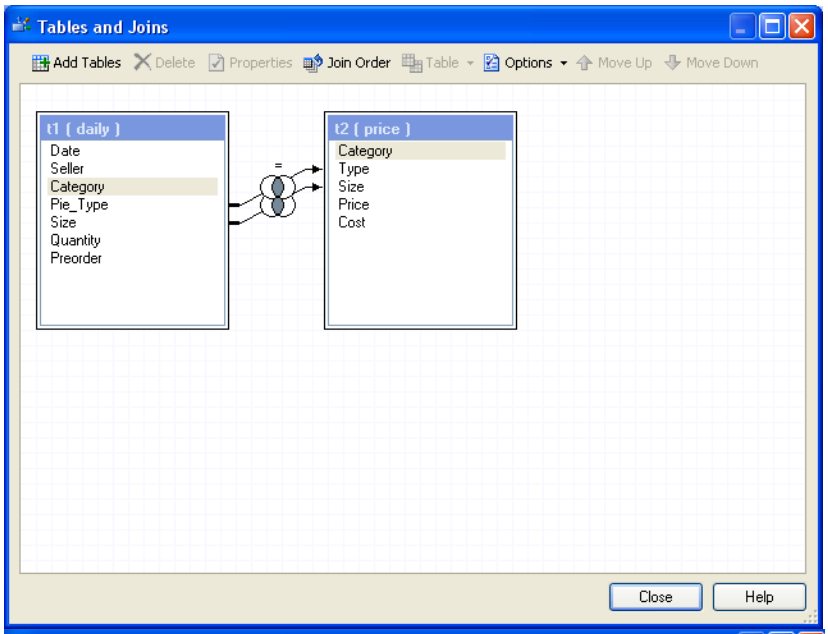


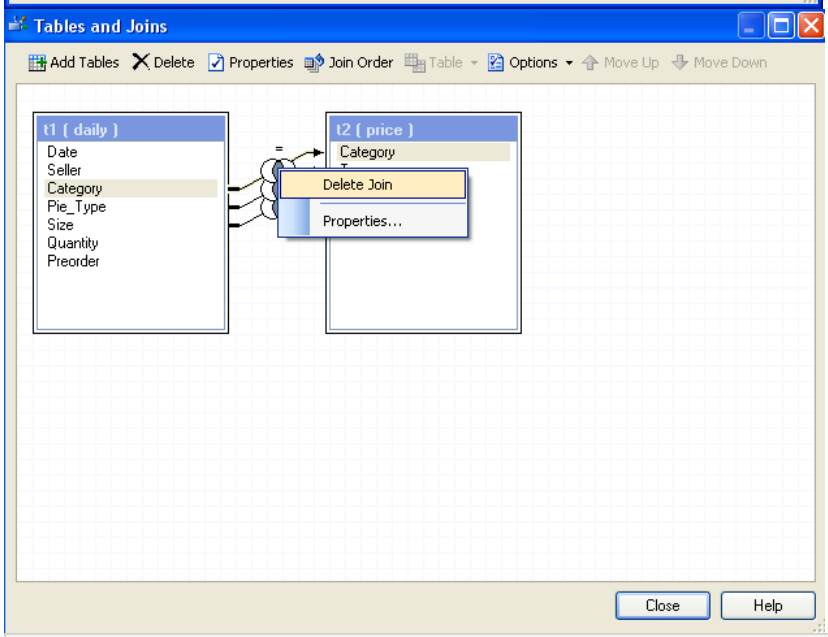
Removing a Join Variable

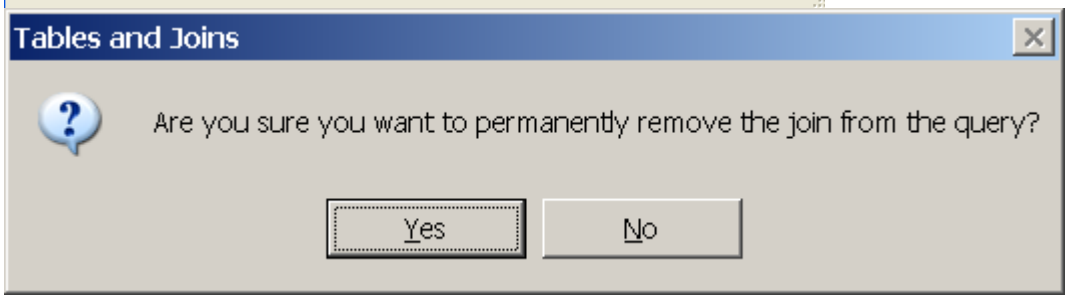
Right-click the join symbol connecting the variables.

Click Delete Join.

Click Yes. Click Close.





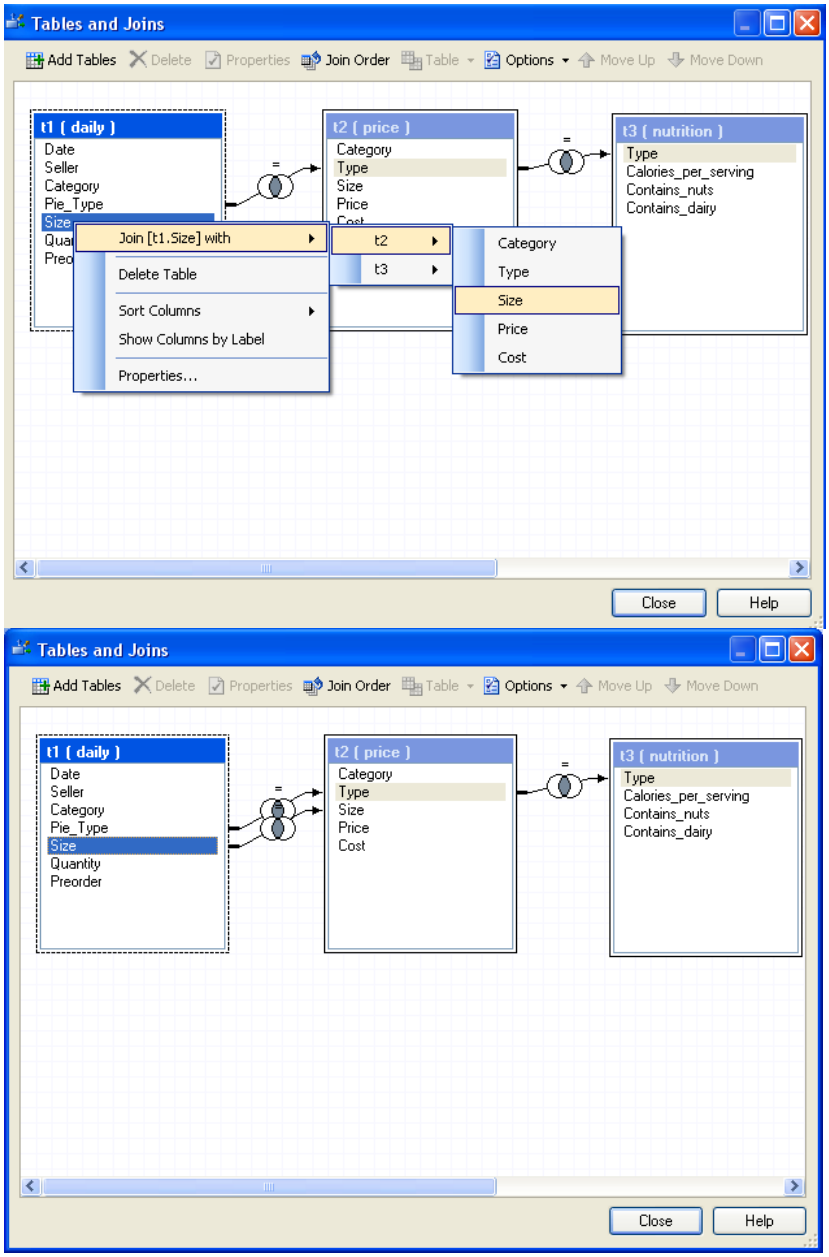


Joining More Than Two Tables

Continue adding data sets.

Modify joins or make manual joins as necessary.

Join up to 32 tables in a single query.



Results (selected columns)

Click Run to run the query.

Date	Category	Pie_Type	Size	Quantity	Price	Cost	Calories per Serving
06/25/2006	Chocola	Columns from Daily data		2	Columns from Price data		Columns from Nutrition data
06/25/2006	Chocolate Pies	Death By Chocolate	8	1			
06/25/2006	Chocolate Pies	Peanut Butter Cup	8	2			
06/25/2006	Chocolate Pies	Peanut Butter Cup	10	1	11.99	6	750
06/25/2006	Chocolate Pies	Black Forest	8	5	7.99	3.95	700
06/25/2006	Fruit Pies	Raspberry	10	2	8.99	4.2	400
06/25/2006	Fruit Pies	Strawberry	10	2	6.99	3.7	485
06/25/2006	Fruit Pies	Blueberry	10	3	7.99	4	375
06/25/2006	Fruit Pies	Lemon Meringue	10	5	10.99	5.65	525
06/25/2006	Fruit Pies	Banana Cream	8	2	7.99	4.2	490
06/25/2006	Fruit Pies	Apple	8	4	6.99	3.7	350
06/25/2006	Fruit Pies	Apple	10	1	8.99	4.5	350

Modifying Joins

By default, all joins are inner joins.

To change the join type:

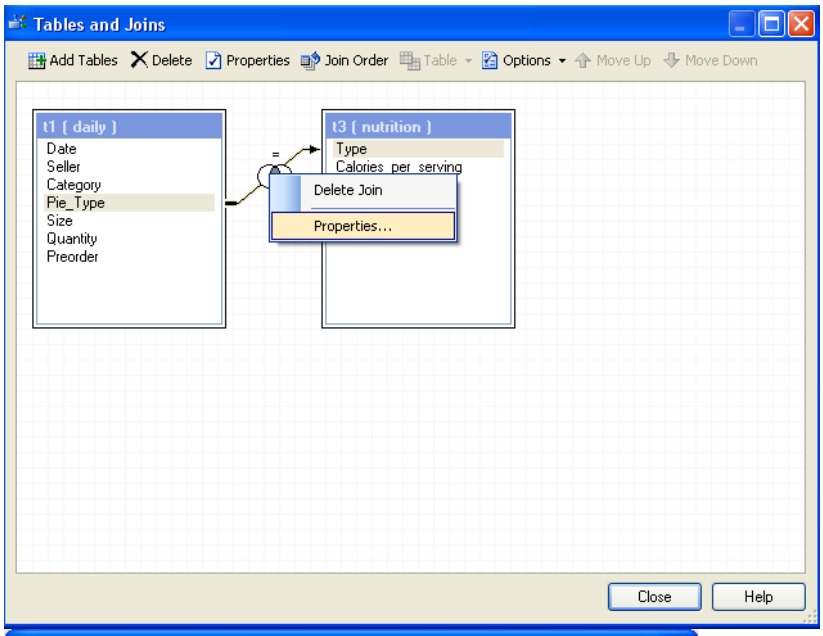
Click the Join Tables button in the Query Builder.

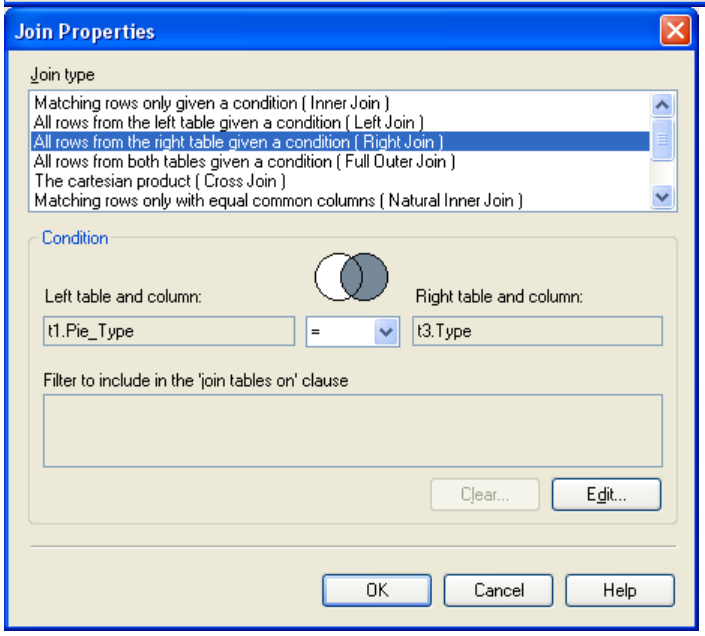
Right-click the join symbol ().

Click Properties.

Under Join type, select the new join type.

Click OK. Click CLOSE.



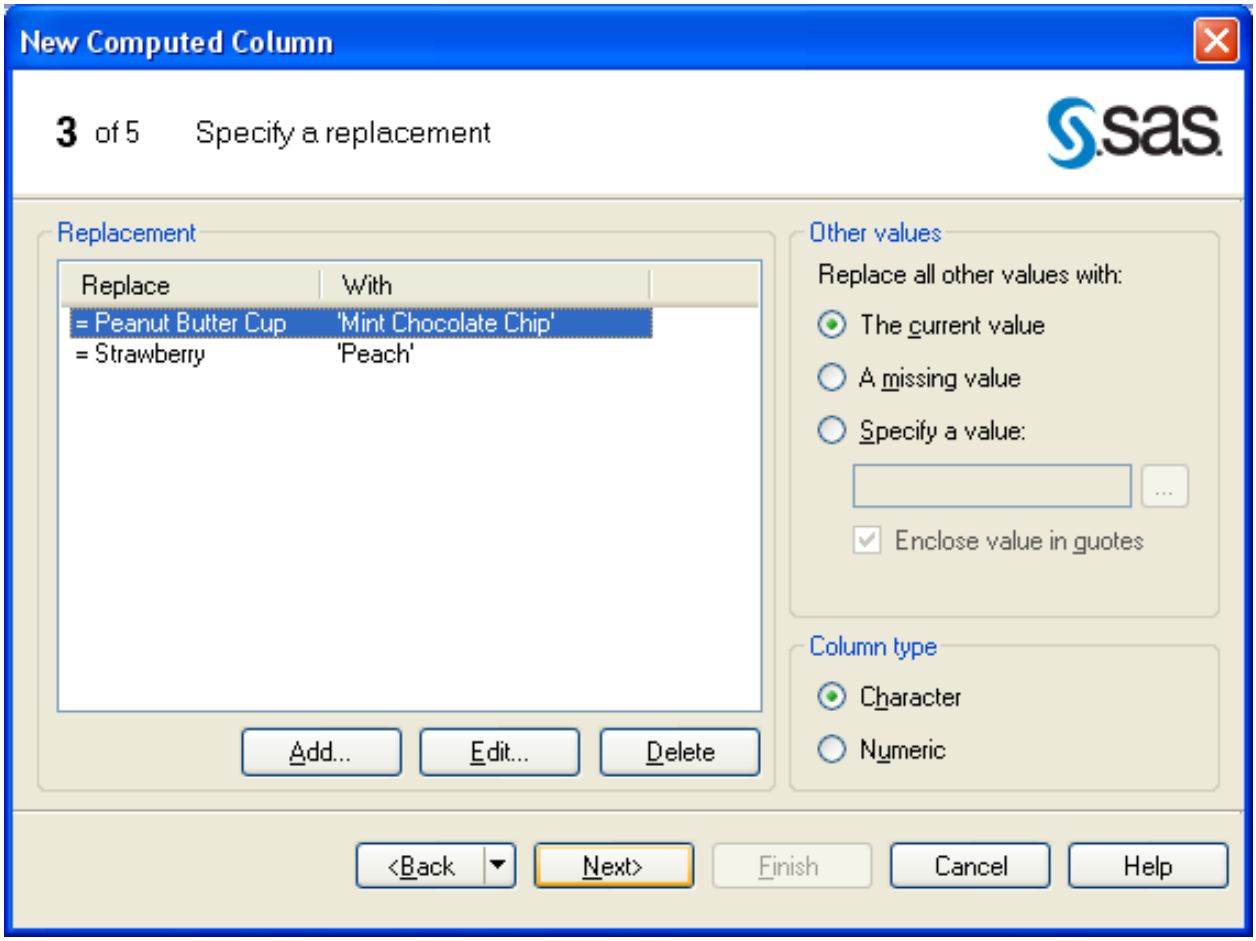


Calculating New Values-Replacement

Replacing values changes select values in a column based on a set of rules you create.

Both discrete values and ranges of values can be replaced.

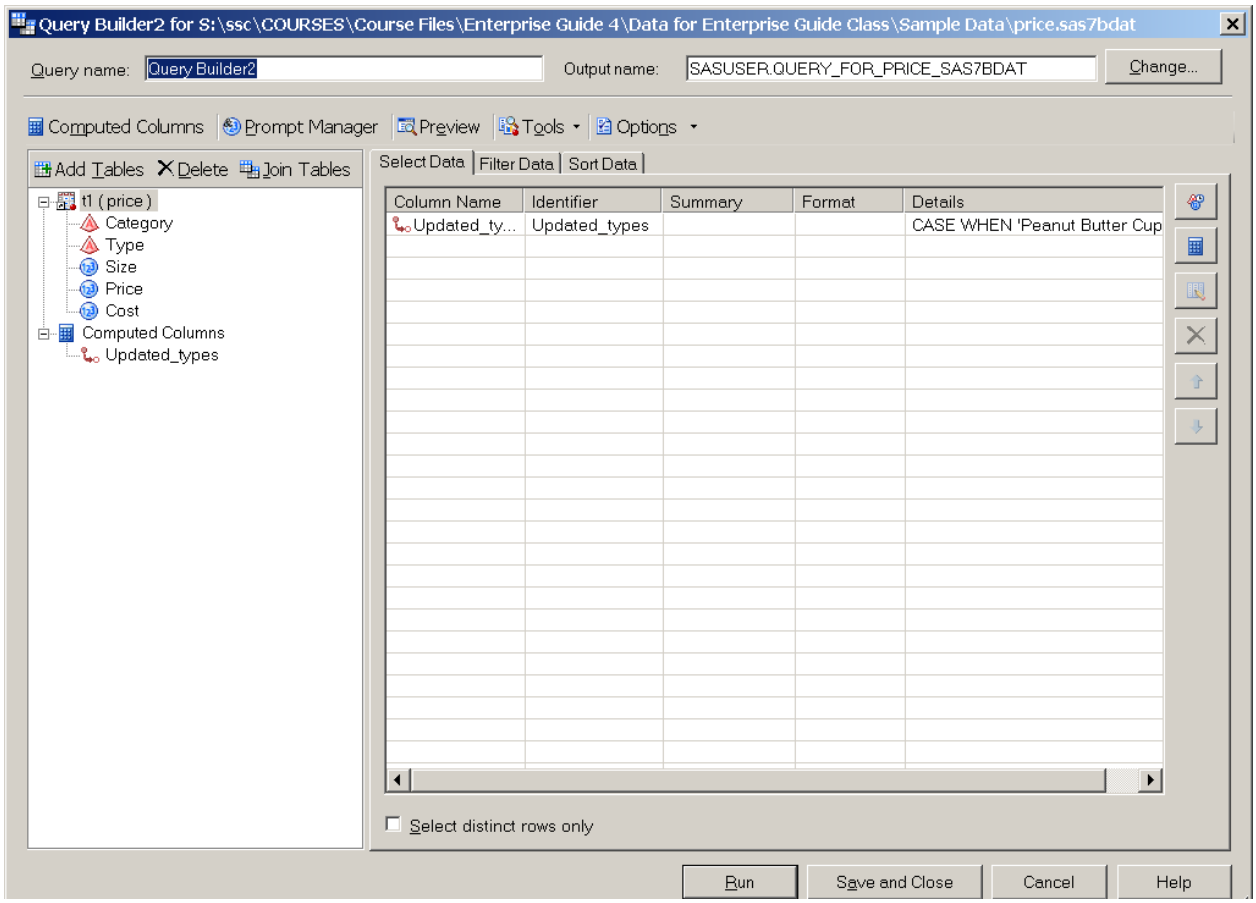
Replacing Discrete Values



Replacing Discrete Values (continued)

The new column will appear on both the Tables list and the Select Data tab.

It can be manipulated in the same ways as other columns in the query.

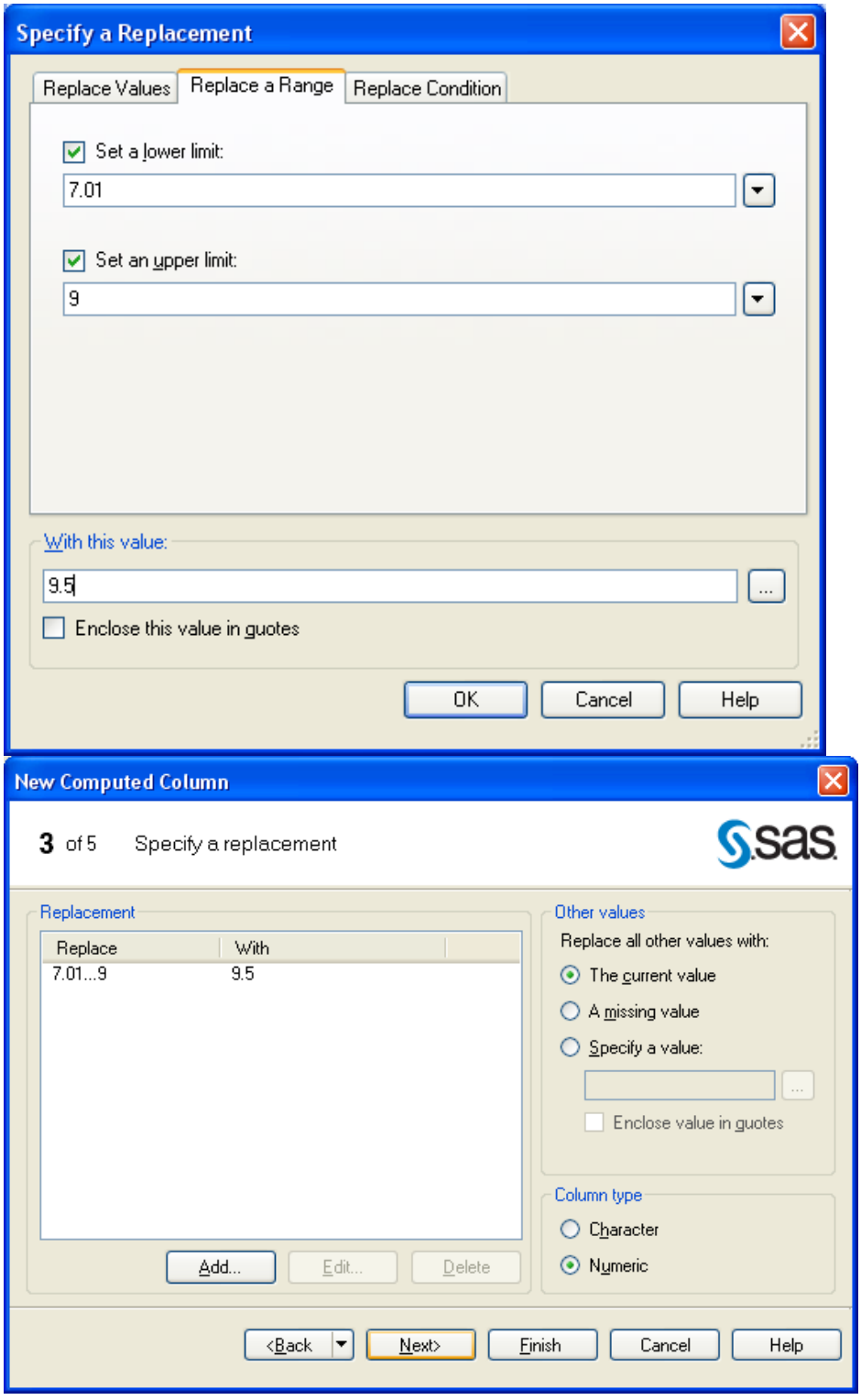


Results

Click Run to run the query.

Category	Type	updated_types	Size	Price	Cost
Chocolate Pies	Chocolate Mousse	Chocolate Mousse	8	8.99	4.25
Chocolate Pies	Chocolate Mousse	Chocolate Mousse	10	10.99	5.5
Chocolate Pies	French Silk	French Silk	8	6.99	3.2
Chocolate Pies	French Silk	French Silk	10	8.99	4.75
Chocolate Pies	Death By Chocolate	Death By Chocolate	8	10.99	5.25
Chocolate Pies	Death By Chocolate	Death By Chocolate	10	12.99	6.85
Chocolate Pies	Peanut Butter Cup	Mint Chocolate Chip	8	9.99	5
Chocolate Pies	Peanut Butter Cup	Mint Chocolate Chip	10	11.99	6
Chocolate Pies	Black Forest	Black Forest	8	7.99	3.95
Chocolate Pies	Black Forest	Black Forest	10	9.99	4.95
Fruit Pies	Raspberry	Raspberry	8	6.99	3.55
Fruit Pies	Raspberry	Raspberry	10	8.99	4.2
Fruit Pies	Rhubarb	Rhubarb	8	5.99	3.15
Fruit Pies	Rhubarb	Rhubarb	10	7.99	3.85
Fruit Pies	Strawberry	Peach	8	4.99	2.6
Fruit Pies	Strawberry	Peach	10	6.99	3.7
Fruit Pies	Blueberry	Blueberry	8	5.99	3.2
Fruit Pies	Blueberry	Blueberry	10	7.99	4
Fruit Pies	Lemon Meringue	Lemon Meringue	8	8.99	4.55
Fruit Pies	Lemon Meringue	Lemon Meringue	10	10.99	5.65
Fruit Pies	Banana Cream	Banana Cream	8	7.99	4.2
Fruit Pies	Banana Cream	Banana Cream	10	9.99	5.1
Fruit Pies	Apple	Apple	8	6.99	3.7
Fruit Pies	Apple	Apple	10	8.99	4.5

Replacing a Range of Values



Results

The new column will appear both on the Tables list and the Select Data tab.

It can be manipulated in the same ways as other columns in the query.

Category	Type	Updated_types	Size	Price	Co	New_Price
Chocolate Pies	Chocolate Mousse	Chocolate Mousse	8	8.99	4.2	9.5
Chocolate Pies	Chocolate Mousse	Chocolate Mousse	10	10.99	5	10.99
Chocolate Pies	French Silk	French Silk	8	6.99	3	6.99
Chocolate Pies	French Silk	French Silk	10	8.99	4.7	9.5
Chocolate Pies	Death By Chocolate	Death By Chocolate	8	10.99	5.2	10.99
Chocolate Pies	Death By Chocolate	Death By Chocolate	10	12.99	6.8	12.99
Chocolate Pies	Peanut Butter Cup	Mint Chocolate Chip	8	9.99		9.99
Chocolate Pies	Peanut Butter Cup	Mint Chocolate Chip	10	11.99		11.99
Chocolate Pies	Black Forest	Black Forest	8	7.99	3.9	9.5
Chocolate Pies	Black Forest	Black Forest	10	9.99	4.9	9.99
Fruit Pies	Raspberry	Raspberry	8	6.99	3.5	6.99
Fruit Pies	Raspberry	Raspberry	10	8.99	4	9.5
Fruit Pies	Rhubarb	Rhubarb	8	5.99	3.1	5.99
Fruit Pies	Rhubarb	Rhubarb	10	7.99	3.8	9.5
Fruit Pies	Strawberry	Peach	8	4.99	2	4.99
Fruit Pies	Strawberry	Peach	10	6.99	3	6.99
Fruit Pies	Blueberry	Blueberry	8	5.99	3	5.99
Fruit Pies	Blueberry	Blueberry	10	7.99		9.5
Fruit Pies	Lemon Meringue	Lemon Meringue	8	8.99	4.5	9.5
Fruit Pies	Lemon Meringue	Lemon Meringue	10	10.99	5.6	10.99
Fruit Pies	Banana Cream	Banana Cream	8	7.99	4	9.5
Fruit Pies	Banana Cream	Banana Cream	10	9.99	5	9.99
Fruit Pies	Apple	Apple	8	6.99	3	6.99

.

Computed Columns

A computed column contains data calculated from values in other columns across each observation.

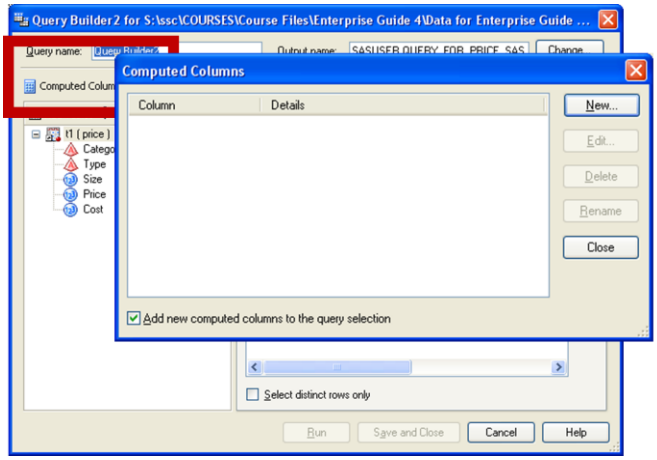
For example, a computed column may contain the sum of two variables.

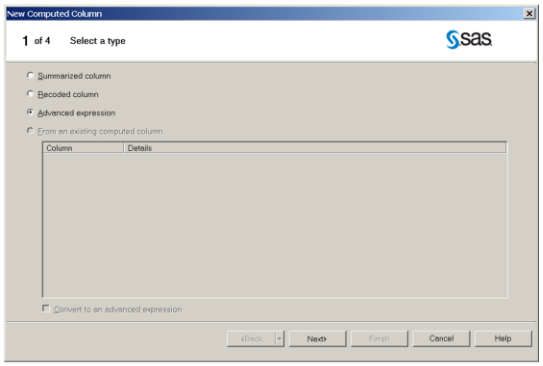
Adding a New Computed Column

In the Query Builder, click the Computed Columns button to open the Computed Columns window.

Click the New button.

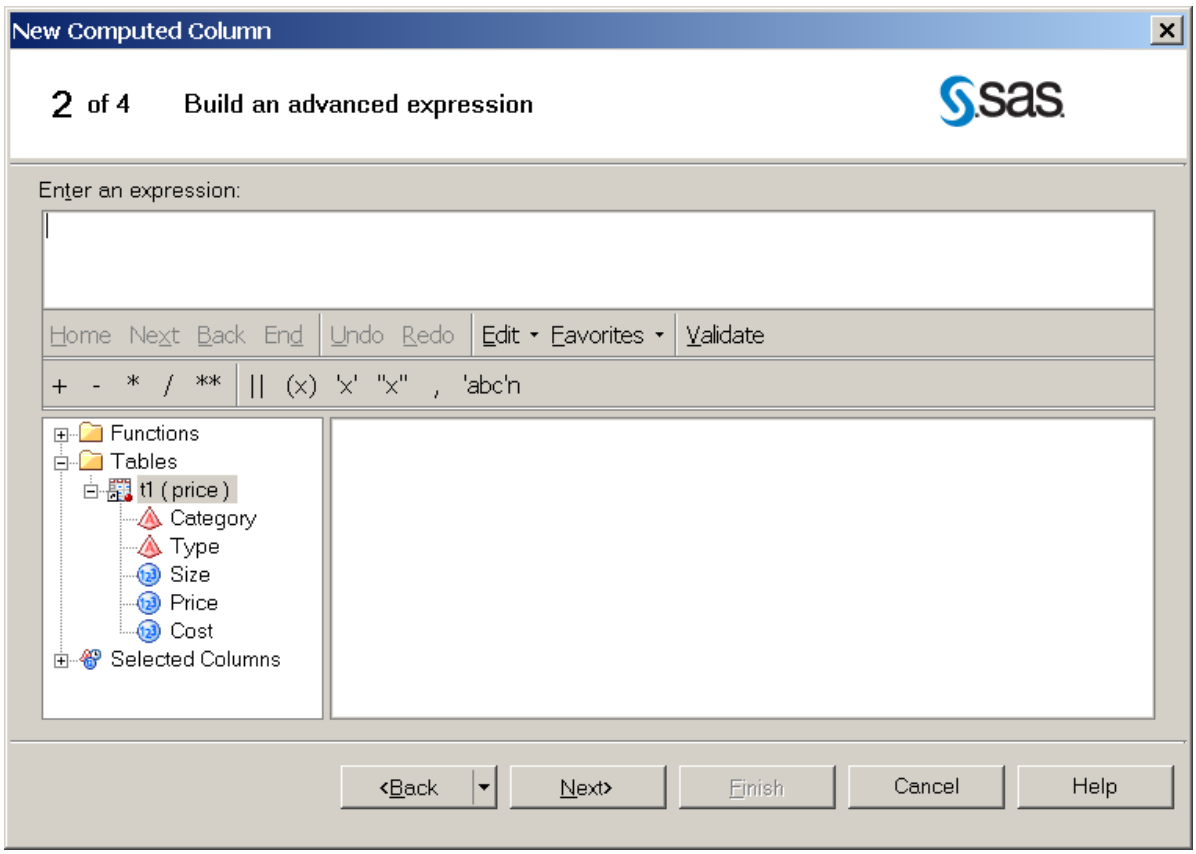
Select Advanced Expression. Click Next.





The Expression Builder

The Expression Builder can be used to create an expression for a new column of data using combinations of operators, variables, numbers and functions.



Building an Expression

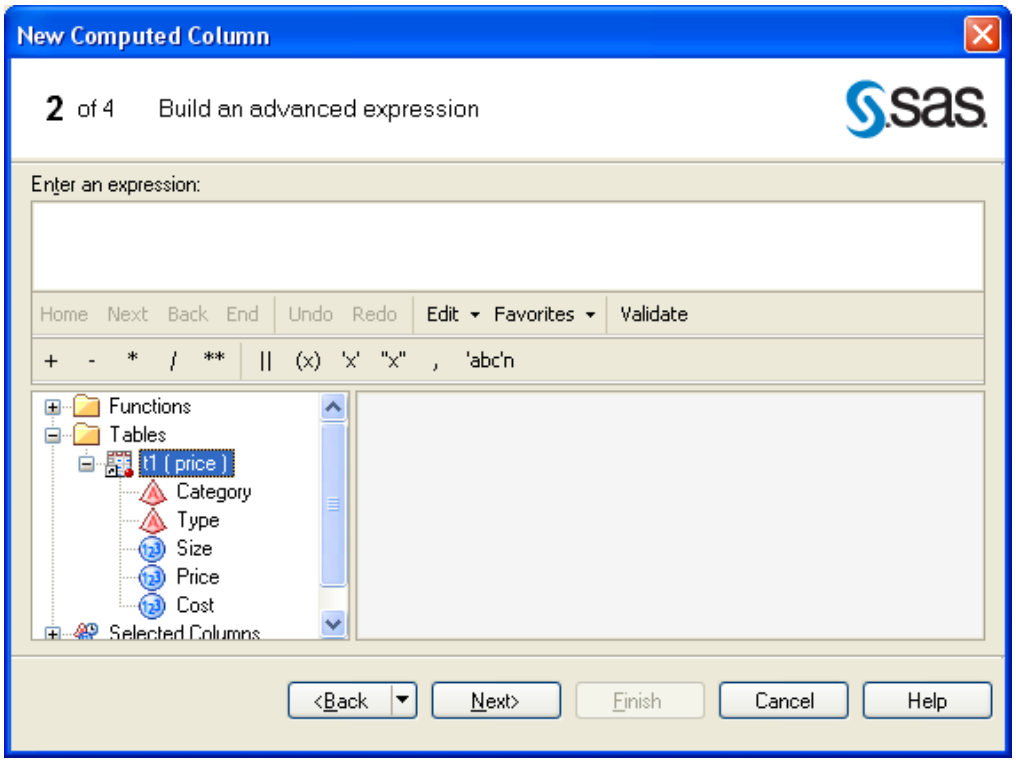
Add elements to the expression box by:

Typing in the Expression Text Box.

Clicking operators on the Operator Bar.

Double-clicking variables on the Variable Values list on the Data tab.

Double-clicking functions on the Functions list on the Functions tab.



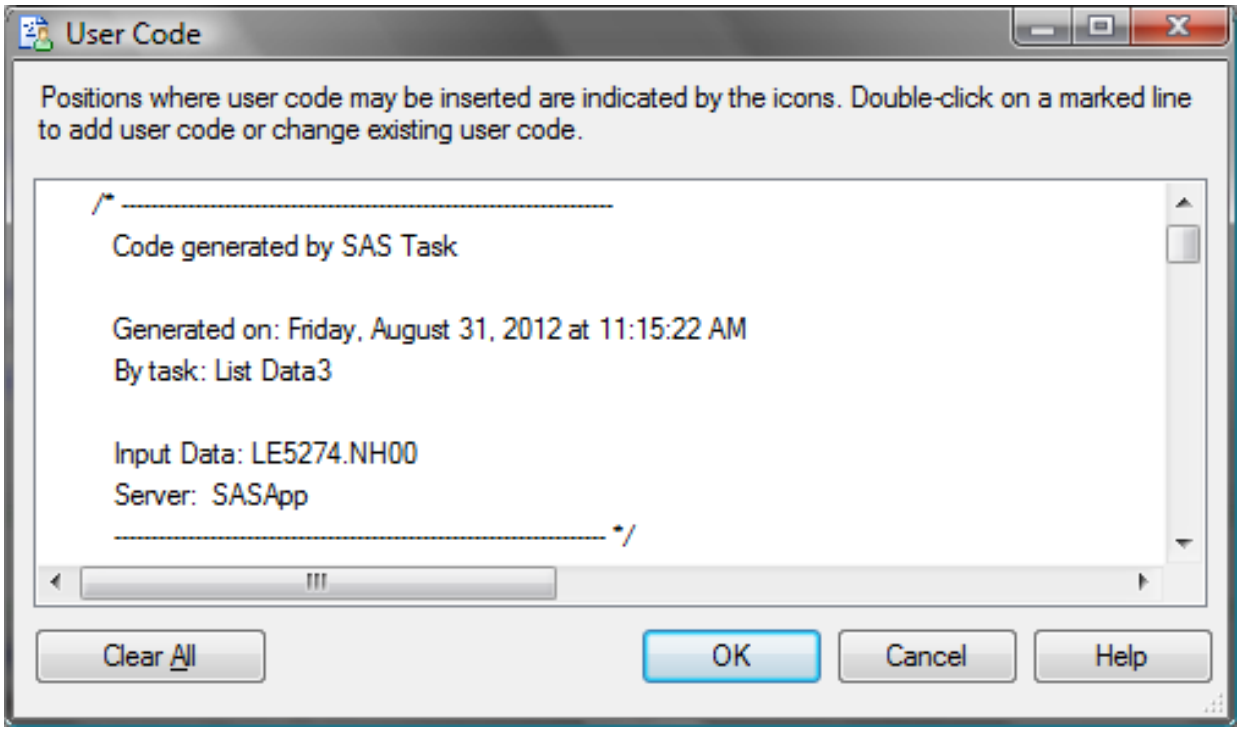
Results

Click Run to run the query.

Category	Type	Size	Price	Cost	Profit
Chocolate Pies	Chocolate Mousse	8	8.99	4.25	4.74
Chocolate Pies	Chocolate Mousse	10	10.99	5.5	5.49
Chocolate Pies	French Silk	8	6.99	3.2	3.79
Chocolate Pies	French Silk	10	8.99	4.75	4.24
Chocolate Pies	Death By Chocolate	8	10.99	5.25	5.74
Chocolate Pies	Death By Chocolate	10	12.99	6.85	6.14
Chocolate Pies	Peanut Butter Cup	8	9.99	5	4.99
Chocolate Pies	Peanut Butter Cup	10	11.99	6	5.99
Chocolate Pies	Black Forest	8	7.99	3.95	4.04
Chocolate Pies	Black Forest	10	9.99	4.95	5.04
Fruit Pies	Raspberry	8	6.99	3.55	3.44
Fruit Pies	Raspberry	10	8.99	4.2	4.79
Fruit Pies	Rhubarb	8	5.99	3.15	2.84
Fruit Pies	Rhubarb	10	7.99	3.85	4.14
Fruit Pies	Strawberry	8	4.99	2.6	2.39
Fruit Pies	Strawberry	10	6.99	3.7	3.29
Fruit Pies	Blueberry	8	5.99	3.2	2.79
Fruit Pies	Blueberry	10	7.99	4	3.99
Fruit Pies	Lemon Meringue	8	8.99	4.55	4.44
Fruit Pies	Lemon Meringue	10	10.99	5.65	5.34
Fruit Pies	Banana Cream	8	7.99	4.2	3.79
Fruit Pies	Banana Cream	10	9.99	5.1	4.89
Fruit Pies	Apple	8	6.99	3.7	3.29
Fruit Pies	Apple	10	8.99	4.5	4.49

Shell Code

You can take the shell SQL code and modify with special options, etc.



The Swiss Army Knife of EG

EG's Query Builder can do so much of what we need in SAS. And it's easy!

Recommended Reading

The Missing Semicolon: http://www.sys-seminar.com/newsletter

Monthly Enterprise Guide Tips

Chris Hemedinger-The SAS Dummy http://blogs.sas.com/content/sasdummy/

Questions, Comments